Mon. Not. R. Astron. Soc. **000**, 000–000 (0000) Printed 4 February 2013 (MN LATEX style file v2.2)

Stability of magnetic fields in non-barotropic stars: an analytic treatment

T. Akgün[1,3⋆], A. Reisenegger[1⋆], A. Mastrano[2] and P. Marchant[1,4]
[1] *Departamento de Astronomía y Astrofísica, Facultad de Física, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, 782-0436 Macul, Santiago, Chile*
[2] *School of Physics, University of Melbourne, Parkville, VIC 3010, Australia*
[3] *Barcelona Supercomputing Center – Centro Nacional de Supercomputación, C/ Gran Capità 2-4, Barcelona, 08034, Spain*
[4] *Argelander Institut für Astronomie, Universität Bonn, Auf dem Hügel 71, D-53121, Bonn, Germany*

4 February 2013

ABSTRACT

Magnetic fields in upper main-sequence stars, white dwarfs, and neutron stars are known to persist for timescales comparable to their lifetimes. From a theoretical perspective this is problematic, as it can be shown that simple magnetic field configurations are always unstable. In non-barotropic stars, stable stratification allows for a much wider range of magnetic field structures than in barotropic stars, and helps stabilize them by making it harder to induce radial displacements. Recent simulations by Braithwaite and collaborators have shown that, in stably stratified stars, random initial magnetic fields evolve into nearly axisymmetric configurations with both poloidal and toroidal components, which then remain stable for some time. It is desirable to provide an analytic study of the stability of such fields. We write an explicit expression for a plausible equilibrium structure of an axially symmetric magnetic field with both poloidal and toroidal components of adjustable strengths, in a non-barotropic static fluid star, and study its stability using the energy principle. We construct a displacement field that should be a reasonable approximation to the most unstable mode of a toroidal field, and confirm Braithwaite's result that a given toroidal field can be stabilized by a poloidal field containing much less energy than the former, as given through the condition $E_{\rm pol}/E_{\rm tor} \gtrsim 2aE_{\rm tor}/E_{\rm grav}$, where $E_{\rm pol}$ and $E_{\rm tor}$ are the energies of the poloidal and toroidal fields, respectively, and $E_{\rm grav}$ is the gravitational binding energy of the star. We find that $a \approx 7.4$ for main-sequence stars, and $a \sim 200$ for neutron stars. Since $E_{\rm pol}/E_{\rm grav} \ll 1$, we conclude that the energy of the toroidal field can be substantially larger than that of the poloidal field, which is consistent with the speculation that the toroidal field is the main reservoir powering magnetar activity.

Key words: instabilities – magnetic fields – MHD – stars: magnetic field – stars: neutron – white dwarfs.

1 INTRODUCTION

Upper main-sequence stars, white dwarfs, and neutron stars are known to possess magnetic fields that persist for long periods of time, comparable to their lifetimes. Since convection does not play an important role in these objects, dynamo generation of magnetic fields is not expected during most of their lives. As a consequence, their magnetic fields must be in stable hydromagnetic equilibrium. However, from a theoretical perspective this poses a problem, since it can be shown that simple magnetic field configurations consisting of purely poloidal (meridional) or purely toroidal (azimuthal) fields are always unstable. In particular, Tayler (1973) showed that toroidal fields are prone to the *interchange* (axisymmetric) and *kink* (non-axisymmetric) instabilities. Markey & Tayler (1973) and Wright (1973) showed similarly that purely poloidal fields, with some field lines closing inside the star, are also unstable near the *neutral line*, where the poloidal field vanishes. Flowers & Ruderman (1977) discussed another large-scale instability of poloidal fields, illustrated by the fact that, when two magnets are aligned, they will tend to orient in opposite direction to one another. Therefore, a rotation of an entire hemisphere of

⋆ E-mail: akgun@astro.cornell.edu (TA); areisene@astro.puc.cl (AR)

a star, cut along a plane containing the axis, should lead to the lowering of the overall energy, as we have demonstrated mathematically (Marchant, Reisenegger & Akgün 2011).

On the other hand, even in the most strongly magnetized stars, the magnetic (Lorentz) force inferred from the surface field strengths is still typically a million times weaker than the hydrostatic force due to pressure and gravity. Therefore, a small perturbation in the non-magnetic background equilibrium could be sufficient to balance the magnetic force. In the radiative envelopes of massive stars and in the interiors of degenerate stars, matter is *non-barotropic* (i.e. pressure depends on a second quantity, such as chemical composition or specific entropy, in addition to density) and *stably stratified*, allowing for a wider range of magnetic field structures than found in *barotropic* fluids (i.e. those where pressure can be expressed as a function of density only) (Reisenegger 2009). Stable stratification also helps stabilize the magnetic field by making it harder to induce radial displacements of the fluid. This effect was included by Tayler (1973); however, by itself it is not sufficient to completely stabilize a purely toroidal (or purely poloidal) magnetic field. Recent simulations for stably stratified stars have demonstrated that initially random magnetic fields tend to evolve into nearly axisymmetric configurations with both poloidal and toroidal components of comparable strength, which then remain stable for several Alfvén times (Braithwaite & Spruit 2004; Braithwaite & Nordlund 2006; Braithwaite 2009). Our goal is to provide an analytic justification for the stability of such fields, and to understand how the poloidal and toroidal components can help stabilize each other.

Before considering the stability, we must first determine the equilibrium structure of the magnetic field. In barotropic stars, the equilibrium form of the magnetic field is severely restricted and is given as the solution of a differential equation (the so-called Grad–Shafranov equation, as discussed, for example, in Chandrasekhar & Fermi 1953; Ferraro 1954; Lüst & Schlüter 1954; Prendergast 1956; a detailed discussion is also given in Akgün & Wasserman 2008). On the other hand, in realistic, non-barotropic stars, the hydrostatic force includes a buoyancy term, which acts as a restoring force for stably stratified fluids. Thus, the only restriction that remains for an axisymmetric field in a non-barotropic fluid is that the magnetic force cannot have an azimuthal ($\hat{\boldsymbol{\phi}}$) component, since no counterpart exists in the hydrostatic force that can act to balance it. In addition, the equilibrium magnetic field needs to satisfy boundary conditions at the surface and regularity conditions at the center of the star. We can construct simple polynomial forms for the scalar functions that describe the poloidal and toroidal components of the magnetic field, consistent with these requirements.

Once we know the equilibrium structure of the magnetic field, we can examine its stability, for which we use the *energy principle* developed by Bernstein et al. (1958). In this method, one considers the energy of perturbations around the magnetic equilibrium. If this energy is positive, then the equilibrium is stable, and vice versa. We then consider the problem of constructing a displacement field that gives rise to instabilities in a purely toroidal field configuration. The hydrostatic and toroidal parts of the energy can be examined analytically for stability, and can be minimized with respect to the azimuthal component of the displacement field, in an analogous manner to Tayler (1973). Once we have found this minimum, we add the poloidal part of the energy and determine how strong the poloidal field must be in comparison to the toroidal field in order to stabilize the field.

The outline of this paper is as follows. In §2 we discuss the equilibrium structure of the star and the magnetic field. We first construct sample equilibrium profiles for the pressure, density, and gravitational potential, which, while being sufficiently simple, have all the desirable qualities. Next, we consider the structure of the poloidal and toroidal fields and discuss their properties. We then construct a simple magnetic field that satisfies the boundary and regularity conditions. In §3 we consider the stability of the magnetic field thus constructed using the energy principle approach. We calculate the contributions to the energy due to the fluid, and due to the poloidal and toroidal components of the magnetic field. We give a proof that all physically relevant, purely toroidal fields are unstable. We discuss the implications of stable stratification on the displacement field. We construct a particular displacement field that makes the sum of the hydrostatic and toroidal parts of the energy negative, yielding an instability, and then show how the addition of a poloidal field eliminates this instability. In §4 we present our conclusions.

2 EQUILIBRIUM

In realistic stars, the stress due to the magnetic field is much weaker than the hydrostatic terms due to pressure and gravity (e.g. Reisenegger 2009). The background equilibrium in the absence of magnetic fields and rotation is spherically symmetric and is given by Euler's equation,

$$\boldsymbol{\nabla} P_0 + \varrho_0 \boldsymbol{\nabla} \Phi_0 = 0 \, , \tag{1}$$

where P is pressure, ϱ is density, and Φ is gravitational potential. Throughout this paper, we will denote the spherically symmetric non-magnetic background quantities with the subscript 0. The gravitational potential is given in terms of the density by Poisson's equation,

$$\nabla^2 \Phi_0 = 4\pi G \varrho_0 \, . \tag{2}$$

The magnetic field $\boldsymbol{B}$ changes the background quantities slightly, and the new equilibrium is given by

$$\boldsymbol{\nabla} P + \varrho \boldsymbol{\nabla} \Phi = \frac{\boldsymbol{J} \times \boldsymbol{B}}{c} \, , \tag{3}$$

where $\boldsymbol{J} = c \boldsymbol{\nabla} \times \boldsymbol{B} / 4\pi$ is the current density. We can express the small changes due to the magnetic field as Eulerian

perturbations, and write $P = P_0 + P_1$, and similarly for ϱ and Φ. Then, we can rewrite the above equation, working to first order in the perturbations, as

$$\nabla P_1 + \varrho_1 \nabla \Phi_0 + \varrho_0 \nabla \Phi_1 = \frac{\boldsymbol{J} \times \boldsymbol{B}}{c} \ . \tag{4}$$

In the often used, idealized assumption of barotropic fluids, there is a unique relation between pressure and density, which holds throughout the application of small perturbations. Therefore, we can write the pressure as a function of the density. This allows us to express the left-hand side of equation (3) (and consequently equation 4) as a gradient of the form $\nabla P + \varrho \nabla \Phi = \varrho \nabla (H + \Phi)$, where $dH(\varrho) = dP(\varrho)/\varrho$. This implies that $\nabla \times (\boldsymbol{J} \times \boldsymbol{B}/\varrho c) = 0$, so the magnetic acceleration must also be expressible as a gradient. This is a strong constraint and greatly restricts the possible choice of the magnetic field in equilibrium (Chandrasekhar & Fermi 1953; Ferraro 1954; Lüst & Schlüter 1954; Prendergast 1956; Akgün & Wasserman 2008; Haskell et al. 2008).

On the other hand, in non-barotropic fluids, pressure depends on at least one additional quantity, as well as density. In white dwarfs and in the radiative zones of non-degenerate stars, the dominant additional quantity is the specific entropy, and in neutron stars it is the composition (fraction of protons or other "impurities"; Reisenegger 2009). For long equilibration times, any changes induced in the background quantities will imply that a simple relation between pressure and density no longer exists. Consequently, the left-hand sides of equations (3) and (4) are not expressible as gradients. Therefore, unlike the barotropic case, we do not require that the magnetic acceleration be expressible as a gradient. Instead, the only constraint for axisymmetric fields is the much less restrictive requirement that the $\hat{\boldsymbol{\phi}}$ component of the magnetic force density vanish, since there is no such component in the hydrostatic part that can balance it (Chandrasekhar & Prendergast 1956; Mestel 1956).

2.1 Non-magnetic equilibrium

In this section, we will consider a simple model for the non-magnetic background equilibrium quantities. The derivations that follow in the subsequent sections do not rely on the specific model, but it will be needed later in the calculation of numerical estimates. A density profile that is simple enough and does not deviate by more than a few percent from an $n = 1$ polytrope (Mastrano et al. 2011), is

$$\varrho_0(x) = \varrho_c (1 - x^2) \ , \tag{5}$$

where ϱ_c is the central density, and we define a dimensionless radial coordinate $x = r/R_\star$, where $R_\star$ is the stellar radius. The mass enclosed within radius x is given by

$$m_0(x) = 4\pi R_\star^3 \int_0^x \varrho_0(x) x^2 dx = \frac{4\pi R_\star^3 \varrho_c}{15} (5x^3 - 3x^5) \ . \tag{6}$$

If the total mass of the star is denoted by $M_\star$, then the central density is $\varrho_c = 15 M_\star / 8\pi R_\star^3$. The gravitational potential inside the star is given by Poisson's equation (equation 2),

$$\Phi_0(x) = \frac{G}{R_\star} \int_0^x \frac{m_0(x)}{x^2} dx = \frac{G M_\star}{8 R_\star} (10x^2 - 3x^4) \ . \tag{7}$$

Here, the gravitational potential at the center is chosen to be zero. From Euler's equation (equation 1), we find that the pressure is given by

$$P_0(x) = P_c - \frac{G}{R_\star} \int_0^x \frac{\varrho_0(x) m_0(x)}{x^2} dx = P_c \left(1 - \frac{5x^2}{2} + 2x^4 - \frac{x^6}{2} \right) \ . \tag{8}$$

The value of the central pressure P_c is determined by requiring that at the surface $P_0(1) = 0$, which yields $P_c = 4\pi G \varrho_c^2 R_\star^2 / 15 = 15 G M_\star^2 / 16\pi R_\star^4$. The profiles of the background quantities P_0, ϱ_0 and Φ_0 are shown in Fig. 1. The gravitational binding energy of the star is

$$E_{\rm grav} = 4\pi G R_\star^2 \int_0^1 x \varrho_0(x) m_0(x) dx = \frac{5 G M_\star^2}{7 R_\star} = \frac{16\pi P_c R_\star^3}{21} \ . \tag{9}$$

2.2 Magnetic field structure

The magnetic field is divergenceless, therefore quite generally it can be expressed as the sum of a *poloidal* and a *toroidal* component, each completely described by a single scalar function (Chandrasekhar 1981). These functions are analogous to the stream functions describing incompressible flows in hydrodynamics. An axisymmetric magnetic field can be written in spherical coordinates (r, θ, ϕ) as

$$\boldsymbol{B} = \boldsymbol{B}_{\rm pol} + \boldsymbol{B}_{\rm tor} = \nabla \alpha(r, \theta) \times \nabla \phi + \beta(r, \theta) \nabla \phi \ . \tag{10}$$

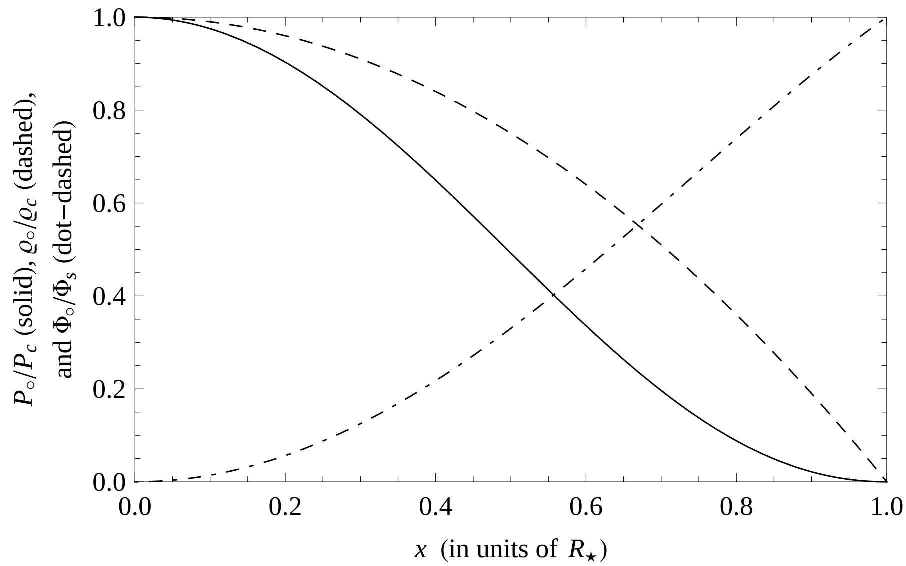

Figure 1. Pressure, density, and gravitational potential profiles chosen for the non-magnetic equilibrium. The pressure and density are scaled by their central values, and the gravitational potential is scaled by its surface value $\Phi_s = 7P_c/4\varrho_c = 7GM_\star/8R_\star$.

Here, we make use of the relation $\boldsymbol{\nabla}\phi = \hat{\boldsymbol{\phi}}/(r\sin\theta)$, which simplifies the calculation of curls. The current densities corresponding to each component are given by

$$\begin{aligned} \frac{4\pi \boldsymbol{J}_{\rm pol}}{c} &= \boldsymbol{\nabla}\times\boldsymbol{B}_{\rm pol} = -\triangle\alpha\boldsymbol{\nabla}\phi \ , \\ \frac{4\pi \boldsymbol{J}_{\rm tor}}{c} &= \boldsymbol{\nabla}\times\boldsymbol{B}_{\rm tor} = \boldsymbol{\nabla}\beta\times\boldsymbol{\nabla}\phi \ . \end{aligned} \tag{11}$$

We have introduced the so-called Grad–Shafranov operator, defining the cylindrical radius as $\varpi = r\sin\theta$,

$$\triangle = \varpi^2\boldsymbol{\nabla}\cdot(\varpi^{-2}\boldsymbol{\nabla}) = \partial_r^2 + \frac{\sin\theta}{r^2}\partial_\theta\left(\frac{\partial_\theta}{\sin\theta}\right) \ . \tag{12}$$

The curl of a poloidal field is a toroidal field, and the curl of a toroidal field is a poloidal field. Therefore, $\boldsymbol{J}_{\rm pol}$ is actually a toroidal field, and $\boldsymbol{J}_{\rm tor}$ is a poloidal field.

We have $\boldsymbol{J}_{\rm pol} \parallel \boldsymbol{B}_{\rm tor} \parallel \hat{\boldsymbol{\phi}}$, therefore the term $\boldsymbol{J}_{\rm pol}\times\boldsymbol{B}_{\rm tor}$ always vanishes in the Lorentz force. On the other hand, for the poloidal components we have $\boldsymbol{J}_{\rm tor}\perp\hat{\boldsymbol{\phi}}$ and $\boldsymbol{B}_{\rm pol}\perp\hat{\boldsymbol{\phi}}$, which implies that $\boldsymbol{J}_{\rm pol}\times\boldsymbol{B}_{\rm pol}\perp\hat{\boldsymbol{\phi}}$ and $\boldsymbol{J}_{\rm tor}\times\boldsymbol{B}_{\rm tor}\perp\hat{\boldsymbol{\phi}}$, while $\boldsymbol{J}_{\rm tor}\times\boldsymbol{B}_{\rm pol}\parallel\hat{\boldsymbol{\phi}}$. However, in axisymmetric equilibrium the Lorentz force cannot have a $\hat{\boldsymbol{\phi}}$ component, as implied by equation (3). Therefore, we must also have $\boldsymbol{J}_{\rm tor}\parallel\boldsymbol{B}_{\rm pol}$, or equivalently $\boldsymbol{\nabla}\alpha\parallel\boldsymbol{\nabla}\beta$, which implies that β can be expressed as a function of α, i.e. $\beta=\beta(\alpha)$. Thus, the Lorentz force can be written as the sum of a term entirely due to the poloidal field, and one entirely due to the toroidal field, $\boldsymbol{f}_{\rm mag} = \boldsymbol{f}_{\rm pol} + \boldsymbol{f}_{\rm tor}$, where

$$\begin{aligned} 4\pi\boldsymbol{f}_{\rm pol} &= (\boldsymbol{\nabla}\times\boldsymbol{B}_{\rm pol})\times\boldsymbol{B}_{\rm pol} = -\varpi^{-2}\triangle\alpha\boldsymbol{\nabla}\alpha \ , \\ 4\pi\boldsymbol{f}_{\rm tor} &= (\boldsymbol{\nabla}\times\boldsymbol{B}_{\rm tor})\times\boldsymbol{B}_{\rm tor} = -\varpi^{-2}\beta\boldsymbol{\nabla}\beta = -\varpi^{-2}\beta\frac{d\beta}{d\alpha}\boldsymbol{\nabla}\alpha \ . \end{aligned} \tag{13}$$

Note that the two terms are poloidal and parallel. Moreover, they are perpendicular to the magnetic surfaces (defined as the surfaces of constant α and β, which contain all the field lines).

2.3 Poloidal field

In this section, we derive a simple profile for an axisymmetric poloidal magnetic field that conforms to certain boundary and regularity conditions. In particular, we impose that there are no surface currents (which would be dissipated very quickly), implying that the poloidal field is continuous across the surface. We assume that the current density drops continuously towards the surface, as the number density of charged particles should be decreasing with the mass density. Moreover, the magnetic field and current density should remain finite and continuous everywhere in the interior, and in particular at the center of the star. We can then construct a polynomial solution for the scalar function that describes the poloidal magnetic field, consistent with these requirements.

Writing the dimensional part of the magnetic field explicitly in terms of some constant $B_{\rm o}$, the poloidal field can be expressed as $\boldsymbol{B}_{\rm pol} = B_{\rm o}\hat{\boldsymbol{\nabla}}\hat{\alpha}\times\hat{\boldsymbol{\nabla}}\phi$ (equation 10). Here, hats denote that the operators are with respect to the dimensionless radial coordinate $x = r/R_\star$, and $\hat{\alpha}$ is also dimensionless. We assume that the field outside the star is that of a point dipole,

which is of the form $\boldsymbol{B}_{\rm dip} \propto x^{-3}(2\hat{\boldsymbol{r}}\cos\theta + \hat{\boldsymbol{\theta}}\sin\theta)$, corresponding to $\hat{\alpha}(x,\theta) \propto \sin^2\theta/x$. In order to match the angular dependence of the field on the surface, we choose

$$\hat{\alpha}(x,\theta) = f(x)\sin^2\theta \ . \tag{14}$$

The poloidal field becomes (equation 10)

$$\boldsymbol{B}_{\rm pol} = B_{\rm o}\hat{\boldsymbol{\nabla}}\hat{\alpha} \times \hat{\boldsymbol{\nabla}}\phi = B_{\rm o}\left[\frac{2f(x)\cos\theta}{x^2}\hat{\boldsymbol{r}} - \frac{f'(x)\sin\theta}{x}\hat{\boldsymbol{\theta}}\right] \ . \tag{15}$$

The current density is (equation 11)

$$\frac{4\pi \boldsymbol{J}_{\rm pol}}{c} = \boldsymbol{\nabla} \times \boldsymbol{B}_{\rm pol} = -\frac{B_{\rm o}}{R_\star}\hat{\triangle}\hat{\alpha}\hat{\boldsymbol{\nabla}}\phi \ , \tag{16}$$

where, from equation (12), we have

$$\hat{\triangle}\hat{\alpha} = \left(f'' - \frac{2f}{x^2}\right)\sin^2\theta \ . \tag{17}$$

Outside the surface, the current density is zero, which implies that

$$f'' = \frac{2f}{x^2} \qquad \text{for} \qquad x > 1 \ . \tag{18}$$

Plugging in a trial solution of the form $f \propto x^s$, we find that the solutions are $s = -1$ and $s = 2$. Thus, outside the star, the solution that remains finite is given by $f \propto x^{-1}$, which is that of a point dipole. (The case $s = 2$ corresponds to a constant magnetic field in the $\hat{\boldsymbol{z}}$ direction.)

Since the density of charged particles decreases to zero at the surface of the star, there cannot be surface currents and the current density has to approach zero at the surface, implying that equation (18) must be satisfied also at $x = 1$. In addition, the magnetic field must be continuous across the surface, which implies that both f and f' should be continuous. Since $f \propto x^{-1}$ outside, it then follows that

$$f' = -\frac{f}{x} \qquad \text{at} \qquad x = 1 \ . \tag{19}$$

Since $f(1) \neq 0$, this equation requires that $|f(x)|$ decrease locally towards the surface. Moreover, we have $|f(0)| = 0 \leqslant |f(1)|$, which implies that $|f(x)|$, or equivalently $|\hat{\alpha}(x,\theta)|$, has at least one maximum somewhere within the star.

In addition to the boundary conditions at the stellar surface (equations 18 and 19), the function f must also satisfy regularity conditions at the center. Since the force density must remain finite, both the magnetic field and the current density must remain finite as well. In particular, for a trial solution of the form $f \propto x^s$, we have $\boldsymbol{B}_{\rm pol} \propto x^{s-2}(2\hat{\boldsymbol{r}}\cos\theta - s\hat{\boldsymbol{\theta}}\sin\theta)$ (equation 15), and $4\pi\boldsymbol{J}_{\rm pol}/c \propto -(s+1)(s-2)x^{s-3}\hat{\boldsymbol{\phi}}\sin\theta$ (equation 16). In order to avoid singularities and multi-valued functions at the origin, we must have either $s = 2$ (corresponding to the zero current case), or $s > 3$. Consistent with this, we seek a solution of the form

$$f(x) = f_2x^2 + f_4x^4 + f_6x^6 \ . \tag{20}$$

We need at least three terms in this polynomial ansatz, in order to be able to satisfy the two homogeneous boundary conditions at the surface (equations 18 and 19). Considering the solution outside the star, and normalizing $f(1) = 1$, we then have

$$f(x) = \begin{cases} \frac{35}{8}x^2 - \frac{21}{4}x^4 + \frac{15}{8}x^6 & \text{for} \quad x \leqslant 1 \ , \\ x^{-1} & \text{for} \quad x > 1 \ . \end{cases} \tag{21}$$

Poloidal field lines are lines of constant $\hat{\alpha}$, and are illustrated in Fig. 2 for the field configuration discussed here. Note that, for a given x, $\hat{\alpha}$ is largest along the equator. It increases smoothly from 0 at the center, reaches a maximum at $x_{\rm max} = \sqrt{(14-\sqrt{21})/15} \approx 0.792$, where its value is $\hat{\alpha}_{\rm max} = f_{\rm max} = (931 + 21\sqrt{21})/900 \approx 1.14$, and then decreases back down to 1 at the surface. The equatorial circle of radius $x_{\rm max}$ is known as the *neutral line*, and the poloidal magnetic field vanishes there. On the other hand, the equation $\hat{\alpha}(x,\theta) = 1$ defines the last magnetic surface that closes within the star. Consequently, the region where $1 \leqslant \hat{\alpha} \leqslant \hat{\alpha}_{\rm max}$ is occupied by field lines closing inside the star.

2.4 Toroidal field

An axisymmetric magnetic field with poloidal and toroidal components can be written as (equation 10)

$$\boldsymbol{B} = B_{\rm o}\left(\eta_{\rm pol}\hat{\boldsymbol{\nabla}}\hat{\alpha} \times \hat{\boldsymbol{\nabla}}\phi + \eta_{\rm tor}\hat{\beta}\hat{\boldsymbol{\nabla}}\phi\right) \ , \tag{22}$$

where $\eta_{\rm pol}$ and $\eta_{\rm tor}$ are dimensionless constants that determine the relative strengths of the two components of the magnetic field. As discussed in §2.2, $\hat{\beta}$ must be expressible as a function of $\hat{\alpha}$. Moreover, the toroidal field must vanish outside the star, since there are no currents to support it there. Since both $\hat{\alpha}$ and $\hat{\beta}$ are constant along the poloidal field lines, it follows that

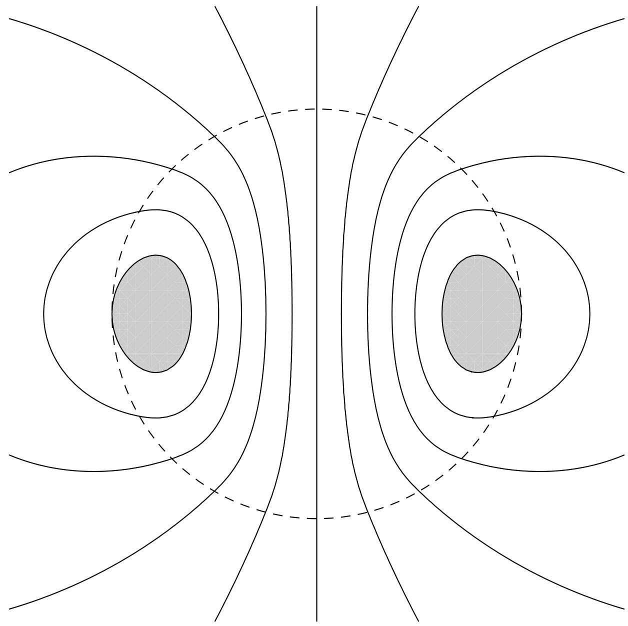

Figure 2. Magnetic field lines for a poloidal field given through equations (14) and (21). α is constant along the field lines. The toroidal field is present only in the shaded donut-shaped region within the star. The stellar surface is shown with a dashed line. The field outside the star is that of a dipole (which is curl-free, i.e. there are no currents outside the star).

the toroidal field is non-zero only in a torus-shaped region defined by the poloidal field lines that close inside the star. The boundary conditions on the poloidal field remain unchanged, and we can still use the results of the previous section (equations 14 and 21). In this case, the last poloidal field line that is closed within the star is given by $\hat{\alpha}(x,\theta) = \hat{\alpha}(1,\pi/2) = 1$. Consider a simple relation of the form

$$\hat{\beta} = \begin{cases} (\hat{\alpha}-1)^n & \text{for} \quad \hat{\alpha} \geqslant 1\,, \\ 0 & \text{for} \quad \hat{\alpha} < 1\,. \end{cases} \tag{23}$$

In order to avoid fast Ohmic dissipation, the current density inside the star must be continuous across the boundary where the toroidal field vanishes. This implies that we must have $n > 1$, so that the current due to the toroidal field decreases smoothly to zero at the boundary. In this paper we will consider the case $n = 2$, but whenever possible we will keep track of the power n for completeness. The equilibrium pressure and density perturbations corresponding to this field structure are calculated in Mastrano et al. (2011).

The extent of the toroidal region in the meridional plane can be calculated from the condition that $\hat{\alpha}(x,\theta) \geqslant 1$. The radial extent is largest along the equator, and its limits are given through the roots of $f(x) = 1$ in the interval $0 \leqslant x \leqslant 1$, which are $x = \sqrt{(27-\sqrt{249})/30} \approx 0.612$ and $x = 1$. The largest angular extent is given by the condition $1/f_{\max} \leqslant \sin^2\theta$, where $f_{\max} \approx 1.14$ as noted in the previous section, which yields $1.21 \lesssim \theta \lesssim 1.93$ in radians (or, $69.4^\circ \lesssim \theta \lesssim 110.6^\circ$). The region where the toroidal field is present is depicted in Fig. 2.

2.5 Amplitude of the magnetic field

The poloidal field is largest along the axis and has a maximum at the origin, while the toroidal field is largest along the equator and has a maximum at $x \approx 0.782$. In our notation, the largest amplitudes of the two components are

$$(B_{\rm pol})_{\max} = \frac{35}{4}\eta_{\rm pol}B_{\rm o} \equiv b_{\rm pol}B_{\rm o} \qquad \text{and} \qquad (B_{\rm tor})_{\max} \approx 0.0254\ \eta_{\rm tor}B_{\rm o} \equiv b_{\rm tor}B_{\rm o}\,. \tag{24}$$

The coefficients $b_{\rm pol}$ and $b_{\rm tor}$ defined in this way are dimensionless. We also note that the surface magnetic field (which is entirely poloidal) has a maximal amplitude of $2\eta_{\rm pol}B_{\rm o}$ at the poles (where it is radial), and a minimal amplitude of $\eta_{\rm pol}B_{\rm o}$ at the equator (where it is tangential to the surface).

Consider the energies stored in the poloidal and toroidal components of the magnetic field,

$$\begin{aligned} E_{\rm pol} &= \frac{1}{8\pi}\int |\boldsymbol{B}_{\rm pol}|^2\, dV = \frac{70}{33}B_{\rm o}^2R_\star^3\eta_{\rm pol}^2 = 2.77\times10^{-2}B_{\rm o}^2R_\star^3b_{\rm pol}^2\,, \\ E_{\rm tor} &= \frac{1}{8\pi}\int |\boldsymbol{B}_{\rm tor}|^2\, dV \approx 4.12\times10^{-6}B_{\rm o}^2R_\star^3\eta_{\rm tor}^2 \approx 6.39\times10^{-3}B_{\rm o}^2R_\star^3b_{\rm tor}^2\,. \end{aligned} \tag{25}$$

The integration for the poloidal part is carried over all of space, while the volume where the toroidal field is present is much smaller.

3 STABILITY

To study the stability of the magnetic field, consider small fluid displacements around the equilibrium given by equation (3),

$$-\varrho\frac{d^2\boldsymbol{\xi}}{dt^2} = \varrho\omega^2\boldsymbol{\xi} = \delta\left(\boldsymbol{\nabla}P + \varrho\boldsymbol{\nabla}\Phi - \boldsymbol{f}_{\rm mag}\right) \equiv -\boldsymbol{\mathcal{F}}(\boldsymbol{\xi})\ . \tag{26}$$

Here, δ denotes Eulerian perturbations due to the displacement field $\boldsymbol{\xi}$, and $\boldsymbol{\mathcal{F}}$ is the net force density induced by the displacements. Note that there are two types of perturbations in our treatment: the magnetically induced ones with respect to the non-magnetic equilibrium, which we denote by the subscript 1 as in equation (4), and those induced by the small displacement $\boldsymbol{\xi}$ with respect to the magnetic equilibrium. The latter can be described either as Eulerian perturbations δ (changes at fixed locations) or Lagrangian perturbations Δ (changes as a fluid element is displaced), which are related through $\Delta = \delta + \boldsymbol{\xi}\cdot\boldsymbol{\nabla}$ (Friedman & Schutz 1978).

There are two ways along which one can proceed from equation (26) in order to determine the stability of the magnetic field configuration. One method is to solve the equation for the perturbations explicitly to determine the frequencies ω, and require them to be all real, $\omega^2 \geqslant 0$. Another method is to employ the energy principle of Bernstein et al. (1958), which has the advantage that one does not need to actually solve the equation; however, it also has the drawback that it is often quite complicated to draw general conclusions. The energy of the perturbations can be written as the sum of hydrostatic and magnetic terms, $\delta W = -\frac{1}{2}\int\boldsymbol{\xi}\cdot\boldsymbol{\mathcal{F}}dV = \delta W_{\rm hyd} + \delta W_{\rm mag}$, where (Akgün & Wasserman 2008)

$$\begin{aligned}\delta W_{\rm hyd} =& \frac{1}{2}\int\left[\Gamma P(\boldsymbol{\nabla}\cdot\boldsymbol{\xi})^2 + (\boldsymbol{\xi}\cdot\boldsymbol{\nabla}P)(\boldsymbol{\nabla}\cdot\boldsymbol{\xi}) - (\boldsymbol{\xi}\cdot\boldsymbol{\nabla}\Phi)(\boldsymbol{\nabla}\cdot\varrho\boldsymbol{\xi}) + \varrho\boldsymbol{\xi}\cdot\boldsymbol{\nabla}\delta\Phi\right]dV \\ &-\frac{1}{2}\oint\left[\Gamma P\boldsymbol{\nabla}\cdot\boldsymbol{\xi} + \boldsymbol{\xi}\cdot\boldsymbol{\nabla}P\right]\boldsymbol{\xi}\cdot d\boldsymbol{S}\ , \\ \delta W_{\rm mag} =& \frac{1}{2}\int\left[\frac{|\delta\boldsymbol{B}|^2}{4\pi} - \frac{\boldsymbol{J}\cdot\delta\boldsymbol{B}\times\boldsymbol{\xi}}{c}\right]dV + \frac{1}{8\pi}\oint\left[\boldsymbol{\xi}(\boldsymbol{B}\cdot\delta\boldsymbol{B}) - \boldsymbol{B}(\boldsymbol{\xi}\cdot\delta\boldsymbol{B})\right]\cdot d\boldsymbol{S}\ .\end{aligned} \tag{27}$$

The magnetic field perturbation follows from Faraday's law of induction,

$$\delta\boldsymbol{B} = \boldsymbol{\nabla}\times(\boldsymbol{\xi}\times\boldsymbol{B})\ . \tag{28}$$

3.1 Implications of non-barotropy

The pressure in a non-barotropic fluid can be written as $P(\varrho, s)$, where s is the specific entropy or chemical composition, depending on the type of star (as discussed in §2). In the non-magnetic background equilibrium, this quantity is a function of density, $s_0(\varrho_0)$, because both s_0 and ϱ_0 are functions of radius. Thus, the background equilibrium is described by a single index,

$$\gamma = \frac{d\ln P_0}{d\ln\varrho_0} = \left(\frac{\partial\ln P}{\partial\ln\varrho}\right)_s + \left(\frac{\partial\ln P}{\partial\ln s}\right)_\varrho\frac{d\ln s_0}{d\ln\varrho_0}\ . \tag{29}$$

For the non-magnetic equilibrium described in §2.1, we have $\gamma(x) = (5-3x^2)/(2-x^2)$, which decreases monotonically from $\gamma(0) = 5/2$ to $\gamma(1) = 2$.

For long equilibration times, the quantity s of a given fluid element remains constant as it is displaced, therefore $\Delta s = 0$. Then, the Lagrangian perturbations of pressure and density are related through

$$\frac{\Delta P}{P} = \left(\frac{\partial\ln P}{\partial\ln\varrho}\right)_s\frac{\Delta\varrho}{\varrho} \equiv \Gamma\frac{\Delta\varrho}{\varrho}\ . \tag{30}$$

Similarly, working to lowest order in B^2 (dropping terms of the order ξB^2), and using $\Delta\varrho = -\varrho\boldsymbol{\nabla}\cdot\boldsymbol{\xi}$, $\delta\varrho = -\boldsymbol{\nabla}\cdot(\varrho\boldsymbol{\xi})$, $\delta s = -\boldsymbol{\xi}\cdot\boldsymbol{\nabla}s \approx -(ds_0/d\varrho_0)\boldsymbol{\xi}\cdot\boldsymbol{\nabla}\varrho_0$, and the definitions of γ and Γ, the Eulerian perturbation of pressure can be written as

$$\frac{\delta P}{P} = \left(\frac{\partial\ln P}{\partial\ln\varrho}\right)_s\frac{\delta\varrho}{\varrho} + \left(\frac{\partial\ln P}{\partial\ln s}\right)_\varrho\frac{\delta s}{s} \approx \gamma\frac{\delta\varrho}{\varrho} + (\Gamma-\gamma)\frac{\Delta\varrho}{\varrho}\ . \tag{31}$$

In a non-barotropic fluid, $\Gamma \neq \gamma$, and the hydrostatic force (which we define as the sum of pressure and gravitational forces, $\boldsymbol{f}_{\rm hyd} = -\boldsymbol{\nabla}P - \varrho\boldsymbol{\nabla}\Phi$) now gives rise to an additional term due to buoyancy, which is proportional to the difference between the indices. Upon the application of small perturbations we have

$$\delta\boldsymbol{f}_{\rm hyd} = -\boldsymbol{\nabla}\delta P - \delta\varrho\boldsymbol{\nabla}\Phi - \varrho\boldsymbol{\nabla}\delta\Phi = -\varrho\boldsymbol{\nabla}\left(\frac{\delta P}{\varrho} + \delta\Phi\right) + \left(\frac{\Gamma}{\gamma} - 1\right)\Delta\varrho\boldsymbol{\nabla}\Phi\ . \tag{32}$$

In a stably stratified star $\Gamma > \gamma$, and the second term acts as a restoring force. Typically, in upper main-sequence stars $\Gamma/\gamma - 1 \sim 1/4$, in white dwarfs $\Gamma/\gamma - 1 \sim T_7/500$, where T_7 is the internal temperature in units of 10^7 K, and in neutron stars $\Gamma/\gamma - 1 \sim$ few % (Reisenegger 2009 and references therein).

3.2 Implications of stable stratification

Consider the integrands of the hydrostatic and magnetic parts given by equation (27). For simplicity, we will always be concerned with cases where the surface integrals vanish (i.e. $\boldsymbol{\xi} = 0$ at the surface). Moreover, we will employ the *Cowling approximation* of neglecting perturbations of the gravitational potential ($\delta\Phi = 0$), which simplifies the calculations considerably. We have

$$\begin{aligned}\mathcal{E}_{\rm hyd} &= \Gamma P(\boldsymbol{\nabla}\cdot\boldsymbol{\xi})^2 + (\boldsymbol{\xi}\cdot\boldsymbol{\nabla}P)(\boldsymbol{\nabla}\cdot\boldsymbol{\xi}) - (\boldsymbol{\xi}\cdot\boldsymbol{\nabla}\Phi)(\boldsymbol{\nabla}\cdot\varrho\boldsymbol{\xi}) \ , \\ \mathcal{E}_{\rm mag} &= \frac{1}{4\pi}\left[|\delta\boldsymbol{B}|^2 - \boldsymbol{\xi}\times(\boldsymbol{\nabla}\times\boldsymbol{B})\cdot\delta\boldsymbol{B}\right] \ .\end{aligned} \tag{33}$$

We can write the adiabatic index of the perturbations as $\Gamma = \Gamma_0 + \Gamma_1$, where, from equation (30),

$$\Gamma_0 = \left(\frac{\partial\ln P}{\partial\ln\varrho}\right)_s\bigg|_{\varrho_0,s_0} \ , \tag{34}$$

and Γ_1 is the magnetic correction. Note that $|\Gamma_1|/\Gamma_0 \sim |P_1|/P_0 \sim |\varrho_1|/\varrho_0 \sim |\Phi_1|/\Phi_0 \sim B^2/P_0 \lesssim 10^{-6}$ (Reisenegger 2009). Thus, the hydrostatic integrand can be rewritten as

$$\begin{aligned}\mathcal{E}_{\rm hyd} =& (\Gamma_0-\gamma)P_0(\boldsymbol{\nabla}\cdot\boldsymbol{\xi})^2 + \frac{\gamma P_0}{\varrho_0^2}(\boldsymbol{\nabla}\cdot\varrho_0\boldsymbol{\xi})^2 \\ &+ (\Gamma_1P_0+\Gamma_0P_1)(\boldsymbol{\nabla}\cdot\boldsymbol{\xi})^2 + (\boldsymbol{\xi}\cdot\boldsymbol{\nabla}P_1)(\boldsymbol{\nabla}\cdot\boldsymbol{\xi}) - (\boldsymbol{\xi}\cdot\boldsymbol{\nabla}\Phi_0)(\boldsymbol{\nabla}\cdot\varrho_1\boldsymbol{\xi}) - (\boldsymbol{\xi}\cdot\boldsymbol{\nabla}\Phi_1)(\boldsymbol{\nabla}\cdot\varrho_0\boldsymbol{\xi}) \ .\end{aligned} \tag{35}$$

For stably stratified stars, $\Gamma_0 > \gamma$, so that the first two terms of the integrand are positive definite. The remaining terms of the integrand are corrections due to the magnetic field. These, as well as $\mathcal{E}_{\rm mag}$, can be positive or negative, but their magnitude is $\lesssim \xi^2B^2/L^2$, where L is some length scale characterizing the spatial variations of the magnetic field. In order for the total energy to be negative, thus allowing for the existence of instabilities, the first two terms of $\mathcal{E}_{\rm hyd}$ must also be small,

$$(\Gamma_0-\gamma)P_0(\boldsymbol{\nabla}\cdot\boldsymbol{\xi})^2 \lesssim \frac{\xi^2B^2}{L^2} \qquad \text{and} \qquad \frac{\gamma P_0}{\varrho_0^2}(\boldsymbol{\nabla}\cdot\varrho_0\boldsymbol{\xi})^2 \lesssim \frac{\xi^2B^2}{L^2} \ . \tag{36}$$

These are constraints that need to be satisfied by the displacement field in order to potentially lead to instabilities. They also imply the following bounds for the remaining terms in $\mathcal{E}_{\rm hyd}$,

$$\begin{aligned}|\Gamma_1|P_0(\boldsymbol{\nabla}\cdot\boldsymbol{\xi})^2 \sim \Gamma_0|P_1|(\boldsymbol{\nabla}\cdot\boldsymbol{\xi})^2 &\lesssim \frac{\Gamma_0\xi^2B^4}{(\Gamma_0-\gamma)P_0L^2} \ , \\ |(\boldsymbol{\xi}\cdot\boldsymbol{\nabla}P_1)(\boldsymbol{\nabla}\cdot\boldsymbol{\xi})| \sim |(\boldsymbol{\xi}\cdot\boldsymbol{\nabla}\Phi_0)(\boldsymbol{\nabla}\cdot\varrho_1\boldsymbol{\xi})| &\lesssim \frac{\xi^2B^3}{\sqrt{(\Gamma_0-\gamma)P_0}L^2} \ , \\ |(\boldsymbol{\xi}\cdot\boldsymbol{\nabla}\Phi_1)(\boldsymbol{\nabla}\cdot\varrho_0\boldsymbol{\xi})| &\lesssim \frac{\xi^2B^3}{\sqrt{\gamma P_0}L^2} \ .\end{aligned} \tag{37}$$

Here, we assume that both γ and Γ_0 are of order unity. Although $\Gamma_0 - \gamma \sim 10^{-2} \ll 1$ in some realistic cases, it is still much larger than the ratio of magnetic pressure to background pressure, $B^2/P_0 \sim 10^{-6}$. Thus, we conclude that (i) corrections to the equilibrium pressure and density due to the magnetic field give rise to terms in the hydrostatic energy that are at least a factor of $B/\sqrt{(\Gamma_0-\gamma)P_0} \lesssim 10^{-2}$ smaller than the (potentially destabilizing) magnetic energy contributions, and therefore can be left out; (ii) the conditions given by equation (36) also imply that the radial component of the displacement field is small, $\xi_r^2/\xi^2 \lesssim B^2/(\Gamma_0-\gamma)P_0 \ll 1$.

3.3 Energy of perturbations for a general displacement field

In this section, we will write down the energy of arbitrary perturbations for poloidal and toroidal fields. Since we assume axisymmetry, and none of the equilibrium quantities depends on the azimuthal angle ϕ, we can express the displacement field in general as a superposition of components of the form

$$\boldsymbol{\xi} = \left[R(r,\theta)\hat{\boldsymbol{r}} + S(r,\theta)\hat{\boldsymbol{\theta}} + iT(r,\theta)\hat{\boldsymbol{\phi}}\right] r\sin\theta e^{im\phi} \ , \tag{38}$$

which can be analyzed separately for different m (as they do not mix in the energy). For notational convenience, we have explicitly written out a factor of cylindrical radius. In general, the dimensionless functions R, S and T will be complex, but only the real part of $\boldsymbol{\xi}$ is physically relevant. Therefore, products should be treated as ZZ^*, where * denotes the complex conjugate.[1]

[1] Caution must be taken in using complex notation to describe real physical quantities. Here, we are dealing with functions of the form $f = F(r,\theta)e^{im\phi}$ and $g = G(r,\theta)e^{im\phi}$, and are interested in integrals of the products of their real parts (denoted by $\Re$), which can be written as

$$\int_0^{2\pi}\Re(f)\Re(g)d\phi = \frac{1}{2}\int_0^{2\pi}\Re(fg^*)d\phi = \pi\Re(FG^*) \ .$$

The energy of the perturbations can be calculated from equation (33). For notational convenience, define an operator Λ and an auxiliary quantity D_m by

$$\Lambda(u) = R\partial_r u + \frac{S\partial_\theta u}{r} \qquad \text{and} \qquad D_m = \frac{\partial_r(r^3 R)}{r^3} + \frac{\partial_\theta(S\sin^2\theta)}{r\sin^2\theta} - \frac{mT}{r\sin\theta} \ . \tag{39}$$

Λ is the directional derivative along the displacement field, $\boldsymbol{\xi}\cdot\boldsymbol{\nabla}u = \Lambda(u) r\sin\theta e^{im\phi}$, where u is an equilibrium quantity independent of the angle ϕ. D_m is the divergence of the displacement field, $\boldsymbol{\nabla}\cdot\boldsymbol{\xi} = D_m r\sin\theta e^{im\phi}$, and we will explicitly keep track of its dependence on m. Defining $\varpi = r\sin\theta$, the hydrostatic part of the integrand (equation 33) becomes

$$\mathcal{E}_{\rm hyd} = \frac{1}{2}\varpi^2\Re\left\{\Gamma P D_m {D_m}^* + [\Lambda(P) - \varrho\Lambda(\Phi)]{D_m}^* - \Lambda(\varrho)\Lambda^*(\Phi)\right\} \ . \tag{40}$$

The factor 1/2 arises as a consequence of the complex notation, as discussed in footnote 1. The equation of hydrostatic equilibrium (equation 3) relates the pressure, density, and gravitational potential to the magnetic field. Therefore, $\mathcal{E}_{\rm hyd}$ depends implicitly on the functions α and β for the poloidal and toroidal components of the magnetic field through the small corrections that these induce on the background quantities. However, as discussed in §3.2, these corrections can be dropped in the calculation of $\mathcal{E}_{\rm hyd}$, so the equilibrium quantities P, ϱ, and Φ in equation (40) can be taken as their non-magnetic versions P_0, ϱ_0, and Φ_0.

The calculation of the magnetic part of the integrand is more involved. The perturbations of the poloidal and toroidal components of the magnetic field (equation 10), are given by equation (28) as

$$\begin{aligned}
\delta\boldsymbol{B}_{\rm pol} =& \boldsymbol{\nabla}\times(\boldsymbol{\xi}\times\boldsymbol{B}_{\rm pol}) = \boldsymbol{\nabla}(\boldsymbol{\xi}\cdot\boldsymbol{\nabla}\phi)\times\boldsymbol{\nabla}\alpha - \boldsymbol{\nabla}(\boldsymbol{\xi}\cdot\boldsymbol{\nabla}\alpha)\times\boldsymbol{\nabla}\phi \\
=& \left\{\frac{mT\partial_\theta\alpha - \partial_\theta[\varpi\Lambda(\alpha)]}{r\varpi}\hat{\boldsymbol{r}} - \frac{mT\partial_r\alpha - \partial_r[\varpi\Lambda(\alpha)]}{\varpi}\hat{\boldsymbol{\theta}} + \frac{i(\partial_r T\partial_\theta\alpha - \partial_\theta T\partial_r\alpha)}{r}\hat{\boldsymbol{\phi}}\right\}e^{im\phi} \ , \\
\delta\boldsymbol{B}_{\rm tor} =& \boldsymbol{\nabla}\times(\boldsymbol{\xi}\times\boldsymbol{B}_{\rm tor}) = (\beta\boldsymbol{\nabla}\phi\cdot\boldsymbol{\nabla})\boldsymbol{\xi} - (\beta\boldsymbol{\nabla}\phi)(\boldsymbol{\nabla}\cdot\boldsymbol{\xi}) - (\boldsymbol{\xi}\cdot\boldsymbol{\nabla})(\beta\boldsymbol{\nabla}\phi) \\
=& \left\{\frac{imR\beta}{\varpi}\hat{\boldsymbol{r}} + \frac{imS\beta}{\varpi}\hat{\boldsymbol{\theta}} - \frac{\partial_r(rR\beta) + \partial_\theta(S\beta)}{r}\hat{\boldsymbol{\phi}}\right\}e^{im\phi} \ .
\end{aligned} \tag{41}$$

Also, using equation (11), we have

$$\begin{aligned}
&\boldsymbol{\xi}\times(\boldsymbol{\nabla}\times\boldsymbol{B}_{\rm pol}) = \boldsymbol{\xi}\times(-\triangle\alpha\boldsymbol{\nabla}\phi) = -\triangle\alpha(S\hat{\boldsymbol{r}} - R\hat{\boldsymbol{\theta}})e^{im\phi} \ , \\
&\boldsymbol{\xi}\times(\boldsymbol{\nabla}\times\boldsymbol{B}_{\rm tor}) = \boldsymbol{\xi}\times(\boldsymbol{\nabla}\beta\times\boldsymbol{\nabla}\phi) = (\boldsymbol{\xi}\cdot\boldsymbol{\nabla}\phi)\boldsymbol{\nabla}\beta - (\boldsymbol{\xi}\cdot\boldsymbol{\nabla}\beta)\boldsymbol{\nabla}\phi = [iT\boldsymbol{\nabla}\beta - \Lambda(\beta)\hat{\boldsymbol{\phi}}]e^{im\phi} \ .
\end{aligned} \tag{42}$$

The magnetic part of the integrand (equation 33) can be written as a sum of three terms: one that is entirely due to the poloidal field, one entirely due to the toroidal field, and a third term that is a combination of the two components, $\mathcal{E}_{\rm mag} = \mathcal{E}_{\rm pol} + \mathcal{E}_{\rm tor} + \mathcal{E}_{\rm cross}$, where

$$\begin{aligned}
\mathcal{E}_{\rm pol} &= \frac{1}{8\pi}\Re\left[\delta\boldsymbol{B}_{\rm pol}\cdot\delta\boldsymbol{B}^*_{\rm pol} - \boldsymbol{\xi}\times(\boldsymbol{\nabla}\times\boldsymbol{B}_{\rm pol})\cdot\delta\boldsymbol{B}^*_{\rm pol}\right] \ , \\
\mathcal{E}_{\rm tor} &= \frac{1}{8\pi}\Re\left[\delta\boldsymbol{B}_{\rm tor}\cdot\delta\boldsymbol{B}^*_{\rm tor} - \boldsymbol{\xi}\times(\boldsymbol{\nabla}\times\boldsymbol{B}_{\rm tor})\cdot\delta\boldsymbol{B}^*_{\rm tor}\right] \ , \\
\mathcal{E}_{\rm cross} &= \frac{1}{8\pi}\Re\left[\delta\boldsymbol{B}_{\rm pol}\cdot\delta\boldsymbol{B}^*_{\rm tor} + \delta\boldsymbol{B}_{\rm tor}\cdot\delta\boldsymbol{B}^*_{\rm pol} - \boldsymbol{\xi}\times(\boldsymbol{\nabla}\times\boldsymbol{B}_{\rm pol})\cdot\delta\boldsymbol{B}^*_{\rm tor} - \boldsymbol{\xi}\times(\boldsymbol{\nabla}\times\boldsymbol{B}_{\rm tor})\cdot\delta\boldsymbol{B}^*_{\rm pol}\right] \ .
\end{aligned} \tag{43}$$

After some algebra, we obtain

$$\begin{aligned}
\mathcal{E}_{\rm pol} =& \frac{1}{8\pi}\left\{\left|\frac{mT\partial_r\alpha - \partial_r[\varpi\Lambda(\alpha)]}{\varpi} + \frac{R\triangle\alpha}{2}\right|^2 + \left|\frac{mT\partial_\theta\alpha - \partial_\theta[\varpi\Lambda(\alpha)]}{r\varpi} + \frac{S\triangle\alpha}{2}\right|^2\right. \\
&\left. + \left|\frac{\partial_r T\partial_\theta\alpha - \partial_\theta T\partial_r\alpha}{r}\right|^2 - \frac{(|R|^2 + |S|^2)(\triangle\alpha)^2}{4}\right\} \ , \\
\mathcal{E}_{\rm tor} =& \frac{1}{8\pi}\left\{\left|\frac{\beta[\partial_r(rR) + \partial_\theta S]}{r} + \frac{\Lambda(\beta)}{2}\right|^2 - \left|\frac{m\beta T}{\varpi} + \frac{\Lambda(\beta)}{2}\right|^2 + \frac{m^2\beta^2(|R|^2 + |S|^2 + |T|^2)}{\varpi^2}\right\} \ , \\
\mathcal{E}_{\rm cross} =& \frac{1}{8\pi}\Re\left\{\frac{iT}{r\varpi}\Big[\partial_r[\varpi\Lambda(\alpha)]\partial_\theta\beta - \partial_\theta[\varpi\Lambda(\alpha)]\partial_r\beta\Big]^* + \frac{2i\beta}{r^2}\Big[\partial_r(rR) + \partial_\theta S\Big]\Big[\partial_r T\partial_\theta\alpha - \partial_\theta T\partial_r\alpha\Big]^*\right. \\
&\left. + \frac{2im\beta}{\varpi}\left[RS^*\triangle\alpha + \frac{mT\partial_r\alpha - \partial_r[\varpi\Lambda(\alpha)]}{\varpi}S^* - \frac{mT\partial_\theta\alpha - \partial_\theta[\varpi\Lambda(\alpha)]}{r\varpi}R^*\right]\right\} \ .
\end{aligned} \tag{44}$$

3.4 Stability of a toroidal field

In this section, following the derivation of Tayler (1973), we consider the problem of constructing a displacement field that makes a purely toroidal magnetic field unstable. In other words, we want to find $\boldsymbol{\xi}$ for which $\mathcal{E}_{\rm hyd} + \mathcal{E}_{\rm tor} < 0$. Then, in the following section we will examine the stability of the poloidal part for the same displacement.

As demonstrated by Tayler (1973) for the purely toroidal field, the real and imaginary parts in the energy separate into

two equivalent terms. Consequently, it is sufficient to consider the case of real R, S, and T. The function T appears only algebraically in the hydrostatic and toroidal parts of the integrand. We then have, from equations (40) and (44),

$$\mathcal{E}_{\rm hyd} + \mathcal{E}_{\rm tor} = \frac{1}{2}\left[E_2(mT)^2 + E_1 mT + E_0\right] \ , \tag{45}$$

where we define,

$$\begin{aligned} E_2 =& \Gamma P \ , \\ E_1 =& -2\varpi\Gamma P D_0 - \varpi\Lambda(P) + \varpi\varrho\Lambda(\Phi) - \frac{\beta\Lambda(\beta)}{4\pi\varpi} \ , \\ E_0 =& \left[\varpi^2\Gamma P + \frac{\beta^2}{4\pi}\right] D_0^2 + \left[\varpi^2\Lambda(P) - \varpi^2\varrho\Lambda(\Phi) + \frac{\beta\Lambda(\beta)}{4\pi} - \frac{\beta^2\Lambda(\varpi)}{\pi\varpi}\right] D_0 \\ & - \varpi^2\Lambda(\varrho)\Lambda(\Phi) - \frac{\beta\Lambda(\beta)\Lambda(\varpi)}{2\pi\varpi} + \frac{\beta^2\Lambda^2(\varpi)}{\pi\varpi^2} + \frac{m^2\beta^2(R^2+S^2)}{4\pi\varpi^2} \ . \end{aligned} \tag{46}$$

Here, $D_0 = D_m + mT/r\sin\theta$ (equation 39) is the only term that contains derivatives of the functions R and S. The above terms can be somewhat simplified using the equation of equilibrium for purely toroidal fields, which follows from equations (3), (13), and (39) as

$$\Lambda(P) + \varrho\Lambda(\Phi) = -\frac{\beta\Lambda(\beta)}{4\pi\varpi^2} \ . \tag{47}$$

Since $E_2 > 0$, the integrand given by equation (45) can be minimized with respect to T for $m \neq 0$. In the minimization, we hold R and S (and therefore D_0) constant. The minimizing value is $mT/\varpi = -E_1/2\varpi E_2 = D_0 - \varrho\Lambda(\Phi)/\Gamma P$ and the minimum of the integrand is $\mathcal{E}_{\rm hyd} + \mathcal{E}_{\rm tor} = E_0/2 - E_1^2/8E_2$. Using equation (39), this minimization corresponds to setting $D_m = \varrho\Lambda(\Phi)/\Gamma P$, which can be alternatively expressed as

$$\Gamma P \boldsymbol{\nabla}\cdot\boldsymbol{\xi} = \varrho\boldsymbol{\xi}\cdot\boldsymbol{\nabla}\Phi \ . \tag{48}$$

Dropping magnetic corrections to the background quantities (which give rise to terms of the order ξB^2), using $\delta\varrho = -\boldsymbol{\nabla}\cdot(\varrho\boldsymbol{\xi})$, $\Delta\varrho = -\varrho\boldsymbol{\nabla}\cdot\boldsymbol{\xi}$, and equation (31), this can be rewritten as

$$\frac{\delta P}{P} \approx \gamma\frac{\delta\varrho}{\varrho} + (\Gamma-\gamma)\frac{\Delta\varrho}{\varrho} \approx 0 \ . \tag{49}$$

Note that $\mathcal{E}_{\rm hyd}$ is a quadratic function of T (equation 40) and $\mathcal{E}_{\rm tor}$ is a linear function of T (equation 44). This implies that both $\mathcal{E}_{\rm hyd}$ and $\mathcal{E}_{\rm hyd}+\mathcal{E}_{\rm tor}$ can be minimized with respect to T for $m \neq 0$. In fact, the minima of the non-magnetic case (which corresponds to minimizing $\mathcal{E}_{\rm hyd}$) and the purely toroidal case (which corresponds to minimizing $\mathcal{E}_{\rm hyd}+\mathcal{E}_{\rm tor}$) are both obtained for the condition given by equation (48). These minima are not precisely identical since the background quantities differ by a small amount between the two cases. Equation (49) implies that the minimum is obtained by setting $\delta P = 0$, which in a barotropic fluid ($\Gamma = \gamma$) further implies that $\delta\varrho = 0$. The minimum of $\mathcal{E}_{\rm hyd}$ to lowest order is $\mathcal{E}_{\rm hyd} = (1/\gamma - 1/\Gamma)(\boldsymbol{\xi}\cdot\boldsymbol{\nabla}P_0)^2/P_0$ (equation 33), which is zero for a barotropic fluid, while for a stably stratified non-barotropic fluid it is positive (as long as $R \neq 0$).

On the other hand, for $m = 0$, we have $\mathcal{E}_{\rm hyd} + \mathcal{E}_{\rm tor} = E_0/2$. Thus, in general, we can combine the two cases ($m = 0$ and $m \neq 0$) and write the energy for any m as

$$\mathcal{E}_{\rm hyd} + \mathcal{E}_{\rm tor} = \frac{E_0}{2} - (1-\delta_{m0})\frac{E_1^2}{8E_2} \ , \qquad \text{where} \qquad \delta_{m0} = \left\{ \begin{array}{lll} 1 & \text{for} & m = 0 \ , \\ 0 & \text{for} & m \neq 0 \ . \end{array}\right. \tag{50}$$

We can further rewrite the integrand by grouping the D_0 terms together and writing them as a complete square, thus separating the derivatives of R and S and leaving out only algebraic terms. Defining $K_m = \delta_{m0}\Gamma P + \beta^2/4\pi\varpi^2$, we have

$$\mathcal{E}_{\rm hyd} + \mathcal{E}_{\rm tor} = \frac{1}{2}\varpi^2 K_m\left\{D_0 - \frac{1}{K_m}\left[\delta_{m0}\varrho\Lambda(\Phi) + \frac{\beta^2\Lambda(\varpi)}{2\pi\varpi^3}\right]\right\}^2 + \frac{1}{2}\varpi^2(a_m R^2 + b_m RS + c_m S^2) \ . \tag{51}$$

Keep in mind that this integrand is already minimized with respect to T for $m \neq 0$. The first term is always positive, and the second term forms a quadratic in R and S. The positive definite term can always be made to vanish by a suitable choice of the displacement field. Therefore, the integrand is always positive if the quadratic is positive, which corresponds to the conditions

$$a_m > 0 \ , \qquad c_m > 0 \qquad \text{and} \qquad b_m^2 < 4a_m c_m \ . \tag{52}$$

These are sufficient and necessary conditions for the *stability* of the toroidal field (Tayler 1973). Note that they are not independent: one of the first two, together with the last one, imply the remaining condition. The coefficients for any m are

given through

$$
\begin{aligned}
a_m =& -\partial_r\varrho\partial_r\Phi - (1-\delta_{m0})\frac{\varrho^2(\partial_r\Phi)^2}{\Gamma P} - \frac{1}{K_m}\left(\delta_{m0}\varrho\partial_r\Phi + \frac{\beta^2}{2\pi r^3\sin^2\theta}\right)^2 \\
&- \frac{\beta\partial_r\beta}{2\pi r^3\sin^2\theta} + \frac{\beta^2}{\pi r^4\sin^2\theta} + \frac{m^2\beta^2}{4\pi r^4\sin^4\theta}\ , \\
b_m =& -\frac{\partial_r\varrho\partial_\theta\Phi}{r} - \frac{\partial_\theta\varrho\partial_r\Phi}{r} - (1-\delta_{m0})\frac{2\varrho^2\partial_r\Phi\partial_\theta\Phi}{r\Gamma P} \\
&- \frac{2}{rK_m}\left(\delta_{m0}\varrho\partial_r\Phi + \frac{\beta^2}{2\pi r^3\sin^2\theta}\right)\left(\delta_{m0}\varrho\partial_\theta\Phi + \frac{\beta^2\cos\theta}{2\pi r^2\sin^3\theta}\right) \\
&- \frac{\beta\partial_r\beta\cos\theta}{2\pi r^3\sin^3\theta} - \frac{\beta\partial_\theta\beta}{2\pi r^4\sin^2\theta} + \frac{2\beta^2\cos\theta}{\pi r^4\sin^3\theta}\ , \\
c_m =& -\frac{\partial_\theta\varrho\partial_\theta\Phi}{r^2} - (1-\delta_{m0})\frac{\varrho^2(\partial_\theta\Phi)^2}{r^2\Gamma P} - \frac{1}{r^2K_m}\left(\delta_{m0}\varrho\partial_\theta\Phi + \frac{\beta^2\cos\theta}{2\pi r^2\sin^3\theta}\right)^2 \\
&- \frac{\beta\partial_\theta\beta\cos\theta}{2\pi r^4\sin^3\theta} + \frac{\beta^2\cos^2\theta}{\pi r^4\sin^4\theta} + \frac{m^2\beta^2}{4\pi r^4\sin^4\theta}\ .
\end{aligned}
\tag{53}
$$

These are equivalent to the results given by Tayler (1973), Goossens & Veugelen (1978), and Akgün & Wasserman (2008), albeit the notation is somewhat different. (Here, we have combined the cases $m = 0$ and $m \neq 0$ into a single general form.) We have $|\varrho\partial_r\Phi| \sim |\Phi\partial_r\varrho| \sim P_0/R_\star$ and $|\varrho\partial_\theta\Phi| \sim |\Phi\partial_\theta\varrho| \sim B^2$, so that, to leading order, the coefficients are

$$
a_m \approx \left(\frac{1}{\gamma} - \frac{1}{\Gamma}\right)\frac{(\partial_r P_0)^2}{P_0} \equiv \varrho_0 N^2 \sim \frac{(\Gamma-\gamma)P_0}{R_\star^2} \qquad \text{and} \qquad |b_m|\ ,\ |c_m| \sim \frac{B^2}{4\pi R_\star^2} \equiv \varrho_0\omega_A^2\ . \tag{54}
$$

Here, N is the Brunt-Väisälä frequency, and ω_A is the Alfvén frequency (i.e. the inverse of the Alfvén crossing-time for the star). In a stably stratified star, $\Gamma > \gamma$, where the two gammas, defined by equations (29) and (30), are of order unity. In this case, the condition $a_m > 0$ is comfortably satisfied, and the problem reduces to showing whether $b_m^2 < 4a_m c_m$ is satisfied, since, if it is true, then the remaining condition $c_m > 0$ follows trivially. However, note that, when $c_m < 0$, the magnetic field is always unstable, immaterial of the value of b_m. The field can also be unstable when c_m is positive, but sufficiently close to zero, while b_m is sufficiently large ($0 < c_m < b_m^2/4a_m \sim B^4/64\pi^2(\Gamma-\gamma)P_0R_\star^2$). This is a very narrow interval. For larger, positive c_m, the condition $b_m^2 < 4a_m c_m$ will always be satisfied.

3.4.1 Coefficients for $m = 0$

For future reference, we quote the coefficients for $m = 0$ here. In this case, we have $K_0 = \Gamma P + \beta^2/4\pi\varpi^2$, and the coefficients given by equation (53) reduce to (Tayler 1973; Goossens & Veugelen 1978; Akgün & Wasserman 2008)

$$
\begin{aligned}
a_0 =& -\partial_r\varrho\partial_r\Phi - \frac{1}{K_0}\left(\varrho\partial_r\Phi + \frac{\beta^2}{2\pi r^3\sin^2\theta}\right)^2 - \frac{\beta\partial_r\beta}{2\pi r^3\sin^2\theta} + \frac{\beta^2}{\pi r^4\sin^2\theta}\ , \\
b_0 =& -\frac{\partial_r\varrho\partial_\theta\Phi}{r} - \frac{\partial_\theta\varrho\partial_r\Phi}{r} - \frac{2}{rK_0}\left(\varrho\partial_r\Phi + \frac{\beta^2}{2\pi r^3\sin^2\theta}\right)\left(\varrho\partial_\theta\Phi + \frac{\beta^2\cos\theta}{2\pi r^2\sin^3\theta}\right) \\
&- \frac{\beta\partial_r\beta\cos\theta}{2\pi r^3\sin^3\theta} - \frac{\beta\partial_\theta\beta}{2\pi r^4\sin^2\theta} + \frac{2\beta^2\cos\theta}{\pi r^4\sin^3\theta}\ , \\
c_0 =& -\frac{\partial_\theta\varrho\partial_\theta\Phi}{r^2} - \frac{1}{r^2K_0}\left(\varrho\partial_\theta\Phi + \frac{\beta^2\cos\theta}{2\pi r^2\sin^3\theta}\right)^2 - \frac{\beta\partial_\theta\beta\cos\theta}{2\pi r^4\sin^3\theta} + \frac{\beta^2\cos^2\theta}{\pi r^4\sin^4\theta}\ .
\end{aligned}
\tag{55}
$$

3.4.2 Coefficients for $m \neq 0$

For $m \neq 0$, the integrand given by equation (51) reduces to

$$
\mathcal{E}_{\rm hyd} + \mathcal{E}_{\rm tor} = \frac{\beta^2}{8\pi r^2}\left[\partial_r(rR) + \partial_\theta S\right]^2 + \frac{1}{2}\varpi^2(a_{m\neq 0}R^2 + b_{m\neq 0}RS + c_{m\neq 0}S^2)\ . \tag{56}
$$

The coefficients are given through (Tayler 1973; Goossens & Veugelen 1978; Akgün & Wasserman 2008)

$$
\begin{aligned}
a_{m\neq 0} &= -\partial_r\varrho\partial_r\Phi - \frac{\varrho^2(\partial_r\Phi)^2}{\Gamma P} - \frac{\beta\partial_r\beta}{2\pi r^3\sin^2\theta} + \frac{m^2\beta^2}{4\pi r^4\sin^4\theta}\ , \\
b_{m\neq 0} &= -\frac{\partial_r\varrho\partial_\theta\Phi}{r} - \frac{\partial_\theta\varrho\partial_r\Phi}{r} - \frac{2\varrho^2\partial_r\Phi\partial_\theta\Phi}{\Gamma P r} - \frac{\beta\partial_r\beta\cos\theta}{2\pi r^3\sin^3\theta} - \frac{\beta\partial_\theta\beta}{2\pi r^4\sin^2\theta}\ , \\
c_{m\neq 0} &= -\frac{\partial_\theta\varrho\partial_\theta\Phi}{r^2} - \frac{\varrho^2(\partial_\theta\Phi)^2}{\Gamma P r^2} - \frac{\beta\partial_\theta\beta\cos\theta}{2\pi r^4\sin^3\theta} + \frac{m^2\beta^2}{4\pi r^4\sin^4\theta}\ .
\end{aligned}
\tag{57}
$$

3.4.3 Proof that all continuous toroidal fields are unstable

Tayler (1973) shows that an instability exists regardless of field strength, if the field has the right structure, and that a toroidal field with a non-zero current density on the axis is necessarily unstable. Goossens, Biront & Tayler (1981) further show that a toroidal field is unstable if there is some point in the star where the field strength is zero ($B = 0$), but the derivative of B with respect to $\sin\theta$ is positive. Next, we show more generally that, in fact, all physically relevant toroidal fields are unstable.

A toroidal field is unstable if, for some value of m, $c_m < 0$ somewhere in the star. For $m = 0$ (equation 55), neglecting perturbations of the order of B^4 caused by the magnetic field, we have

$$c_0 = -\frac{\beta\partial_\theta\beta\cos\theta}{2\pi r^4\sin^3\theta} + \frac{\beta^2\cos^2\theta}{\pi r^4\sin^4\theta} = -\frac{\sin\theta\cos\theta}{4\pi r^4}\partial_\theta\left(\frac{\beta^2}{\sin^4\theta}\right) . \tag{58}$$

On the other hand, for $m \neq 0$, we have, from equation (57),

$$c_{m\neq 0} = -\frac{\beta\partial_\theta\beta\cos\theta}{2\pi r^4\sin^3\theta} + \frac{m^2\beta^2}{4\pi r^4\sin^4\theta} = -\frac{\tan^{m^2-1}\theta\,\partial_\theta(\beta^2\cot^{m^2}\theta)}{4\pi r^4\sin^2\theta} . \tag{59}$$

Considering specifically $m = 1$, this simplifies to

$$c_1 = -\frac{\partial_\theta(\beta^2\cot\theta)}{4\pi r^4\sin^2\theta} . \tag{60}$$

Now, we need to look at the behavior of β. In this case, the magnetic field and current density are $\boldsymbol{B} = \beta\boldsymbol{\nabla}\phi$ and $4\pi\boldsymbol{J}/c = \boldsymbol{\nabla}\beta\times\boldsymbol{\nabla}\phi$, respectively (from equations 10 and 11). At the axis, the magnetic field and current density cannot have $\hat{\boldsymbol{\varpi}}$ and $\hat{\boldsymbol{\phi}}$ components, which implies that β must go to zero faster than $\varpi \propto \sin\theta$. Then, as $\varpi \to 0$, we have $4\pi\boldsymbol{J}/c \to \varpi^{-1}\partial_\varpi\beta\hat{\boldsymbol{z}}$. If we want the latter to be finite, we need β to go to zero at least as fast as $\varpi^2 \propto \sin^2\theta$ as we approach the axis. Thus, in the above equations (for $m = 0$ and $m \neq 0$), β^2 easily cancels the singularities due to $\sin\theta$ at $\theta = 0$ and π.

In particular, consider the coefficient c_0 as given by equation (58). If $\beta \propto \sin^2\theta$ then c_0 vanishes everywhere, which, in the best case, implies marginal stability (if $b_0 = 0$ as well). If, on the other hand, β goes to zero faster than $\sin^2\theta$ (as is the case for the field considered in this paper, for which $\beta = 0$ identically in a finite range of θ), then the function $\beta^2/\sin^4\theta$ increases from zero to a finite value somewhere in the interval $0 < \theta < \pi/2$ (i.e. it has a positive derivative while $\cos\theta > 0$), and decreases from some finite value to zero for $\pi/2 < \theta < \pi$ (i.e. it has a negative derivative while $\cos\theta < 0$). Therefore, in both cases we will have some regions where $c_0 < 0$, thus leading to instability.

On the other hand, as can be seen from equation (60), we will have $c_1 < 0$ whenever the derivative $\partial_\theta(\beta^2\cot\theta)$ is positive. Note that $\beta^2\cot\theta = 0$ at $\theta = 0$, $\pi/2$, and π; $\beta^2\cot\theta \geqslant 0$ for $0 < \theta < \pi/2$, and $\beta^2\cot\theta \leqslant 0$ for $\pi/2 < \theta < \pi$. Thus, if β is a continuous function, we will have $c_1 < 0$ somewhere in whichever hemisphere β has some non-zero values. In other words, the $m = 1$ instabilities will happen in those regions where, moving on a spherical shell of constant radius in the direction of increasing θ, $\beta^2\cot\theta$ increases from zero to some finite value while $\cot\theta > 0$, or from some finite negative value back to zero while $\cot\theta < 0$.

3.4.4 Application to our particular magnetic field structure

Now consider the application to our choice of toroidal magnetic field, where β is given by equation (23), and α is given by equations (14) and (21). Keep in mind the renormalization of the functions α and β carried out in accordance with equation (22) (we are now considering the case $\eta_{\rm pol} = 0$ and $\eta_{\rm tor} = 1$, so we have only a toroidal field, despite taking $\alpha \neq 0$). For completeness, in the next few lines, we will keep track of the power n defined in equation (23). Thus, $\alpha(x,\theta) = f(x)\sin^2\theta$ and $\beta(x,\theta) = [\alpha(x,\theta) - 1]^n$ in the region where the toroidal field is non-zero. Note that this β vanishes long before reaching the axis. We consider only the region where $\beta \neq 0$, i.e. $\alpha > 1$, or equivalently, $1/f(x) < \sin^2\theta \leqslant 1$. From equations (58) and (60), we have

$$\frac{c_{0,1}}{B_\circ^2/4\pi R_\star^2} = -\frac{[f(x)\sin^2\theta - 1]^{2n-1}}{x^4\sin^4\theta}G_{0,1} , \tag{61}$$

where $G_0(x,\theta) = 4\cos^2\theta[(n-1)f(x)\sin^2\theta + 1]$ and $G_1(x,\theta) = f(x)\sin^2\theta(4n\cos^2\theta - 1) + 1$. In equation (61), the coefficient of $G_{0,1}$ is always non-positive, and the instability condition $c_{0,1} < 0$ thus requires $G_{0,1} > 0$. Note that for our particular choice of the magnetic field $G_0 \geqslant 0$, thus $c_0 \leqslant 0$, i.e. there is always an $m = 0$ instability. For $m = 1$, the instability condition $G_1 > 0$ is satisfied for $\sin^2\theta \approx 1/f(x)$, but is not satisfied for $\sin^2\theta \approx 1$, i.e. c_1 changes sign somewhere between these two values. The point where the sign change takes place is given as the real root of $G_1 = 0$,

$$\sin^2\theta_{\rm c} = \frac{1}{2}\left[1 - \frac{1}{4n} + \sqrt{\left(1 - \frac{1}{4n}\right)^2 + \frac{1}{nf(x)}}\right] . \tag{62}$$

The unstable region is $1/f(x) < \sin^2\theta < \sin^2\theta_{\rm c}$, and the stable region is $\sin^2\theta_{\rm c} < \sin^2\theta \leqslant 1$. For $n = 2$ and $f(x) = f_{\max} \approx 1.14$ (§2.3), corresponding to the largest extent in colatitude, the stable region is $83.2^\circ < \theta < 96.8^\circ$, and the rest of the interval where the toroidal field is present, $69.4^\circ < \theta < 110.6^\circ$ (§2.4), is unstable. The contours of the coefficients c_0 and c_1 are shown in Fig. 3.

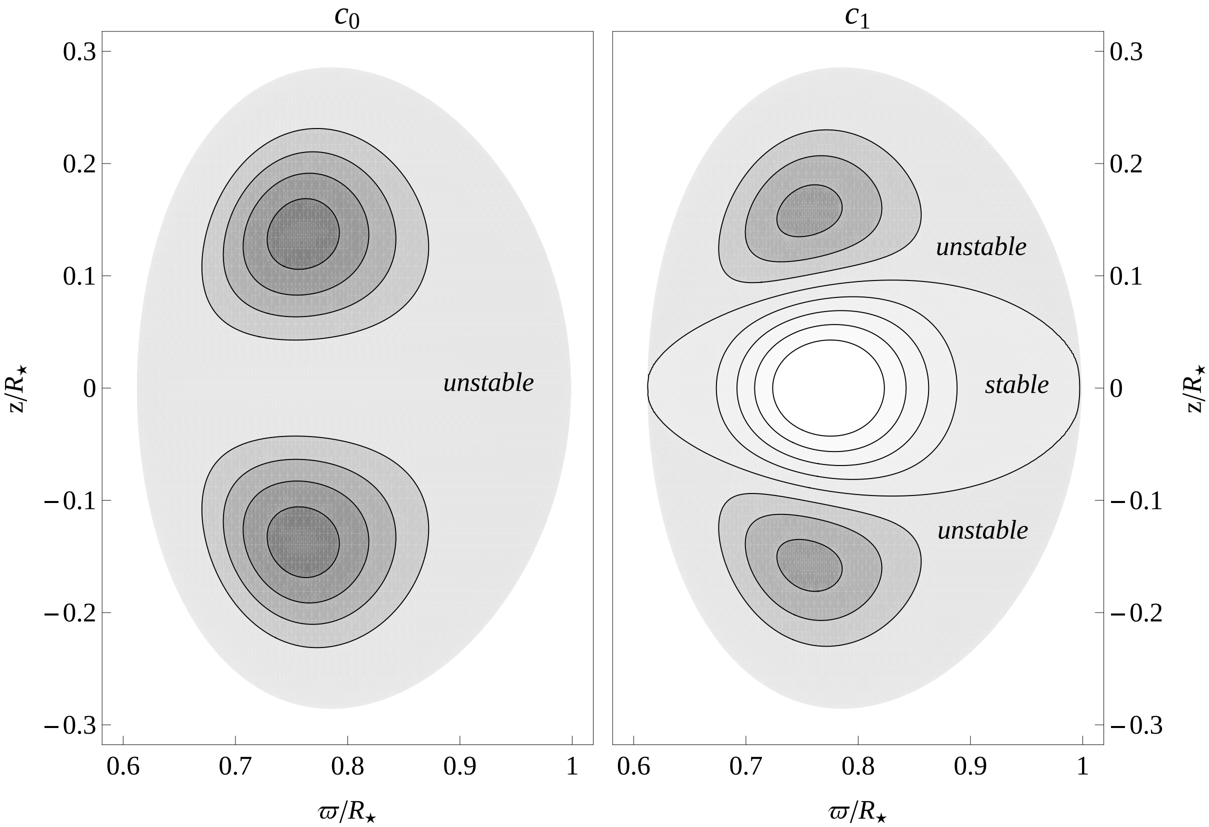

Figure 3. Contours of Tayler's coefficients c_0 (equation 55) and c_1 (equation 57). The coefficients are shown in units of $b_{\rm tor}^2 B_{\rm o}^2/4\pi R_\star^2$. The equator (horizontal) and axis (vertical) are shown in units of stellar radii ($\varpi/R_\star$ and $z/R_\star$, respectively). The outer boundary is defined by the last poloidal field line that closes within the star (and is tangential to the surface at the equator); both $\beta = 0$ and $c_m = 0$ on this boundary. This defines the region where the toroidal field exists. The toroidal field is of the form given by equation (23) with $n = 2$. The contours are shown in the range -1.2 (dark) to -0.3 (light) for c_0, and -0.9 (darkest) to 1.2 (white) for c_1 in increments of 0.3. $c_m < 0$ implies instability. c_0 is zero along the equator ($\theta = \pi/2$) and on the boundary, and is negative everywhere else. The maximum of c_1 occurs along the equator, at $x \approx 0.772$; its minima are at $x \approx 0.772$ and $\theta \approx \pi/2 \pm 0.205$; its radial extent along the equator is $0.612 \lesssim x \leqslant 1$; and its angular extent in the meridional plane is $1.21 \lesssim \theta \lesssim 1.93$.

3.4.5 Proof that the limiting case of perfect stable stratification implies stability

Typically, the hydrostatic force is much stronger than the magnetic force. This implies that, in a stably stratified star, any radial displacement will be acting against a prohibitively large buoyancy force. In the limiting case of *perfect stable stratification*, let's consider a displacement field that is perpendicular to the restoring hydrostatic force, which to lowest order points in the radial direction (equation 32). In addition, we require the fluid to be incompressible. In other words, the density remains constant as a fluid element is displaced (i.e. $\Delta\varrho = -\varrho\boldsymbol{\nabla}\cdot\boldsymbol{\xi} = 0$, while $\Delta P \neq 0$, implying that $\Gamma \to \infty$ from equation 30). These two conditions, namely *incompressibility* ($\boldsymbol{\nabla}\cdot\boldsymbol{\xi} = 0$) and *orthogonality to the radial direction* ($\hat{\boldsymbol{r}}\cdot\boldsymbol{\xi} = 0$), imply that the displacement field is described by a single unknown function (instead of three, as is the case for an unrestricted vector field). From these assumptions it also follows that the hydrostatic part of the energy vanishes (to first order in B^2, as in equation 35). Therefore, all we are left with is the variation in the magnetic energy. In our notation, the requirements $\boldsymbol{\nabla}\cdot\boldsymbol{\xi} = 0$ and $\hat{\boldsymbol{r}}\cdot\boldsymbol{\xi} = 0$ correspond to setting $D_m = 0$ and $R = 0$ in the integrand of equation (45). To first order in B^2, the integrand reduces to (for any m)

$$\mathcal{E}_{\rm hyd} + \mathcal{E}_{\rm tor} = \frac{1}{8\pi r^2}\left[\beta^2(\partial_\theta S)^2 + \frac{(m^2\beta^2 - \beta\partial_\theta\beta\sin 2\theta)S^2}{\sin^2\theta}\right] . \tag{63}$$

This displacement field is restricted to such an extent that it is not possible to make the first positive definite term vanish by a suitable choice, unlike in the general case. Therefore, the coefficient of S^2 is no longer sufficient to assess stability. In fact, it is possible to show that the integral is always positive, which is not immediately obvious from the above form. It is made clearer by rewriting the integrand as

$$\mathcal{E}_{\rm hyd} + \mathcal{E}_{\rm tor} = \frac{1}{8\pi r^2}\left[\frac{(m^2-1)S^2\beta^2}{\sin^2\theta} + \beta^2(\partial_\theta S + S\cot\theta)^2 - \frac{\partial_\theta(S^2\beta^2\cos\theta)}{\sin\theta}\right] . \tag{64}$$

The last term integrates to zero, since $S\sin\theta \to 0$ and $\beta/\sin\theta \to 0$ on the symmetry axis. On the other hand, the sum of the first two terms is always positive for $m^2 \geqslant 1$. For $m = 0$, the only displacement field consistent with $\boldsymbol{\nabla}\cdot\boldsymbol{\xi} = 0$ and $\hat{\boldsymbol{r}}\cdot\boldsymbol{\xi} = 0$

that does not diverge is of the form $\boldsymbol{\xi} = \xi_\phi \hat{\boldsymbol{\phi}}$, which has no effect on the toroidal field (equation 28). Thus, we conclude that the equilibrium is always stable to this very restricted set of perturbations, as previously noted by Dicke (1979). Therefore, in order to obtain instabilities, the restrictions due to perfect stable stratification must be relaxed.

3.4.6 Constraints on the destabilizing displacement field

In this section we will consider the properties of the displacement field that destabilizes the toroidal magnetic field. Some simple observations can be inferred by noting that the energy integrand can be written as the sum of a positive definite term and a quadratic as in equation (51). We would like the quadratic to be negative and the positive definite term to be as small as possible.

First, consider the quadratic, $Q(R,S) \equiv a_m R^2 + b_m RS + c_m S^2$. Since a_m is always large and positive, and c_m is small and negative in some region, $|R|$ must be small compared to $|S|$ in order to allow the energy to become negative. On the other hand, R cannot be zero, as that reduces the integrand in equation (56) to the form given by equation (63), which was shown to be always positive. This is because the integrand in equation (56) is already minimized with respect to T, i.e. we have implicitly substituted the condition given by equation (48), which implies that if $R = 0$ then $D_m \approx 0$, thus leading us back to equation (63) for the perfect stable stratification. Thus, $|R|$ *must be small compared to* $|S|$*, but non-zero.* Since $a_m > 0$, $Q(R,S)$ can be minimized with respect to R. The minimum is given by

$$R_{\min} = -\frac{b_m}{2a_m} S \qquad \text{and} \qquad Q_{\min} = \left(-\frac{b_m^2}{4a_m} + c_m\right) S^2 \ . \tag{65}$$

For instability, we must have $Q_{\min} < 0$, implying that $b_m^2 > 4a_m c_m$. Since typically $|b_m| \sim |c_m| \ll a_m$ (equation 54), this will be satisfied when $c_m < 0$ (plus a thin region where $c_m \geqslant 0$ but very small), and we will have $Q_{\min} \approx c_m S^2$. If $c_m < 0$ is confined to a region of size $(\Delta r, \Delta\theta)$ in the relevant coordinates, the displacement field should also be roughly confined to this region, as there would otherwise be a positive contribution to the energy from the region where $c_m > 0$. Thus, in particular, $|\partial_\theta S| \gtrsim |S|/\Delta\theta$ (since S must vanish near the boundary of the region, we have $|\Delta S| \sim |S|$).

The positive definite terms for $m = 0$ and $m \neq 0$ are significantly different, and the two cases need to be treated separately. First, we will consider the case $m \neq 0$. Since the region we are interested in is near the equator (i.e. $\sin\theta \approx 1$ and $\cos\theta \sim \Delta\theta$), we have $c_{m\neq 0} \sim -\beta^2/4\pi r^4$ (equation 60). For instability, we need $\mathcal{E}_{\rm hyd} + \mathcal{E}_{\rm tor} < 0$ in equation (56), which implies $[\partial_r(rR) + \partial_\theta S]^2 \lesssim S^2$. Thus, we need $|\partial_r(rR) + \partial_\theta S| \lesssim |S| \lesssim |\partial_\theta S|\Delta\theta$, using the above inequality. In other words, the more confined the displacement field is in latitude (i.e. the smaller the range $\Delta\theta$, forced by the condition $c_{m\neq 0} < 0$), the more precisely the two derivatives on the left-hand side need to cancel each other. In the limit $\Delta\theta \to 0$, it is necessary to enforce

$$\partial_r(rR) + \partial_\theta S = 0 \ . \tag{66}$$

In addition, from the confinement to an interval Δr we have $|\partial_r(rR)| \gtrsim r|R|/\Delta r$, and from the cancelation with $\partial_\theta S$ we have $|\partial_r(rR)| \gtrsim |S|/\Delta\theta$. The ratio of these two lower bounds is, using equations (54) and (65), $\mathfrak{R} = (|R|/|S|)(r\Delta\theta/\Delta r) \approx (|b_{m\neq 0}|/2a_{m\neq 0})(r\Delta\theta/\Delta r) \sim (B^2/8\pi P_0)(r\Delta\theta/\Delta r) \ll 1$. Thus, the lower bound on $|\partial_r(rR)|$ from the requirement of canceling $\partial_\theta S$ (even if not exactly) is much larger than the bound from being confined to the region $(\Delta r, \Delta\theta)$. Thus, the length scale of variation of R must be $\delta_r \ll \Delta r$. Note that this does not mean that $\boldsymbol{\xi}$ is confined to a region as thin as δ_r. It could extend over the whole Δr, but it would have to oscillate on a radial length scale δ_r.

For $m = 0$ the coefficient of the positive definite term in equation (51) is much larger than in the $m \neq 0$ case, $K_0 \gg K_{m\neq 0}$. This implies that the quantity in parentheses must cancel out even more precisely. Keeping only leading order terms, we get

$$\frac{\partial_r(r^3 R)}{r^3} + \frac{\partial_\theta(S\sin^2\theta)}{r\sin^2\theta} + \frac{\partial_r P_0}{\Gamma P_0} R = 0 \ . \tag{67}$$

This is different from equation (66) for the $m \neq 0$ case. In what follows we will consider only the simpler case of $m \neq 0$, and the $m = 0$ case will be left for future work.

3.4.7 Particular displacement field for $m \neq 0$

As discussed in the previous section, the quadratic part of the integrand for $m \neq 0$ (equation 56) can be made negative by a suitable choice of the amplitudes of the functions R and S. In addition, the first term, which is positive definite, can be minimized by a suitable choice of the derivatives of these functions. In particular, the best choice might be when this term is made to vanish (Goossens & Tayler 1980; Goossens & Biront 1980), which leads to the condition given by equation (66). This equation is satisfied by solutions of the form

$$R = \frac{\partial_\theta \Pi}{x} \qquad \text{and} \qquad S = -\partial_x \Pi \ , \tag{68}$$

where x is the dimensionless radial coordinate $x = r/R_\star$, and $\Pi(x,\theta)$ is some scalar generating function for the displacement field. While we can choose R and S so that the positive definite term vanishes, T has a particular value for which the integrand in equation (45) is minimized. This value of T is expressible in terms of R, S, and their derivatives, corresponding to the condition given by equation (48). Using equation equations (39) and (66), and keeping only the lowest order terms, we have

$$\frac{mT}{r\sin\theta} = D_0 - \frac{\varrho\Lambda(\Phi)}{\Gamma P} \approx \left(\frac{2}{r} + \frac{\partial_r P_0}{\Gamma P_0}\right) R + \left(\frac{2\cot\theta}{r}\right) S \ . \tag{69}$$

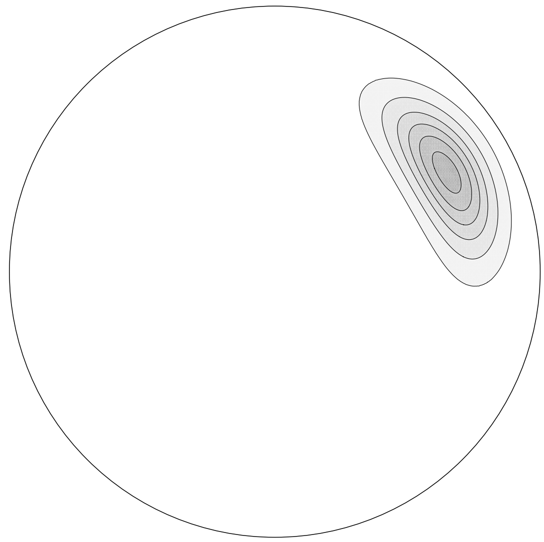

Figure 4. Contours of a greatly exaggerated sample generating function Π for the displacement field (equation 72). The contours are the streamlines of the displacement field. The stellar surface is shown as a solid line. The values of the various parameters used in this plot are $x_0 = 3/4$, $\theta_0 = \pi/3$, $\delta_r = 1/5$, $\delta_\theta = \pi/5$, and $\sigma = 3$. The actual generating function used in our calculations is much more confined.

As noted in §3.2, to lowest order, we can drop all magnetic corrections to the background quantities in the hydrostatic part of the energy (equation 40), which then becomes, using the equation of equilibrium (equation 1), the definition of γ (equation 29), and the value of D_m from equation (48),

$$\mathcal{E}_{\rm hyd} \approx \frac{1}{2}\left(\frac{1}{\gamma} - \frac{1}{\Gamma}\right)\frac{(\partial_r P_0)^2}{P_0} R^2 r^2 \sin^2\theta \approx \frac{1}{2} a_{m\neq 0} R^2 r^2 \sin^2\theta \ . \tag{70}$$

The last equality follows from equation (54). In a stably stratified star $\Gamma > \gamma$, so that the hydrostatic part is always positive. The toroidal part of the energy follows from equation (44), to leading order and using $|R| \ll |S|$,

$$\mathcal{E}_{\rm tor} \approx \frac{1}{2} c_{m\neq 0} S^2 r^2 \sin^2\theta \ . \tag{71}$$

Thus, the total integrand given by equation (56) reduces to $\mathcal{E}_{\rm hyd} + \mathcal{E}_{\rm tor} \approx \frac{1}{2}(a_{m\neq 0}R^2 + c_{m\neq 0}S^2) r^2 \sin^2\theta$. Since $|R| \ll |S|$ and $a_{m\neq 0} \gg |b_{m\neq 0}| \sim |c_{m\neq 0}|$ (equation 54), it follows that the $b_{m\neq 0}RS$ term in the quadratic can be dropped.

We can now proceed to construct a particular displacement field that will make the purely toroidal magnetic field unstable. We will assume that the displacement field is confined to a region within the star and is zero everywhere else. In order to prove that the stability conditions are both sufficient and necessary, Tayler (1973) assumes a particular solution of the form $\Pi(x,\theta) \propto \sin kx \sin \ell\theta$ in a finite volume, bounded by a surface on which $\Pi(x,\theta) = 0$. This corresponds to a finite displacement field which is tangential to the boundaries. While this form is acceptable for a purely toroidal field (since both $\boldsymbol{\xi}$ and $\boldsymbol{B}$ are tangential to the surface), in our case we will eventually incorporate a poloidal field as well, and any discontinuity in the displacement field at the boundaries would cause divergences (cutting the field lines). In order to avoid such pathologies, we would therefore like $\boldsymbol{\xi}$ to go to zero at the boundary, and its derivatives to remain finite everywhere. Thus, we choose

$$\Pi(x,\theta) = \frac{\xi_{\rm o}}{R_\star}\left[1 - \chi^2(x,\theta)\right]^\sigma \ , \qquad \text{where} \qquad \chi^2(x,\theta) = \frac{(x-x_0)^2}{\delta_r^2} + \frac{(\theta-\theta_0)^2}{\delta_\theta^2} \ . \tag{72}$$

The factor $\xi_{\rm o}/R_\star$ sets the amplitude of the displacement field (equation 38). The displacement field is zero on the boundary (defined by $\chi = 1$) and outside of it. This particular choice of χ corresponds to a donut-shaped region with a meridional cross-section in the shape of a distorted ellipse (Fig. 4). In order for the derivatives of the displacement field to remain finite, we must have $\sigma \geqslant 2$.

As shown in §3.4.3, the second condition in equation (52) is always violated somewhere for any non-singular toroidal magnetic field. Therefore, we will have $c_{m\neq 0} < 0$ in some region (Fig. 3). If we choose the displacement field to be confined near the minimum of $c_{m\neq 0}$ and make $c_{m\neq 0}S^2$ to be the dominant term in the quadratic in equation (56), then the energy will be negative. Thus, we want, roughly, $a_{m\neq 0}R^2 \lesssim |c_{m\neq 0}|S^2$. For a displacement field given by equations (68) and (72), this implies that we must have $|R|/|S| \sim \delta_r/\delta_\theta \lesssim \sqrt{|c_{m\neq 0}|/a_{m\neq 0}} \sim B/\sqrt{P_0} \ll 1$. Here, we have used $|x - x_0| \leqslant \delta_r$, $|\theta - \theta_0| \leqslant \delta_\theta$, and equation (54). Since δ_θ cannot be much larger than the angular extent of the negative region of the coefficient $c_{m\neq 0}$ (Fig. 3), this then imposes a very stringent upper limit on δ_r.

3.5 Stability of a toroidal field in the presence of a weaker poloidal component

The displacement field constructed in the previous section makes the sum of the hydrostatic and toroidal parts of the energy negative, thus leading to an instability. In the present section we will consider the case when a weaker poloidal component is added. This poloidal field will give an additional positive contribution to the energy and will help stabilize the instability of the toroidal field. Our goal is to determine the minimum strength of the poloidal field relative to the toroidal field in order to achieve stability. Note that this treatment inherently relies on the implicit assumption (validated by the eventual results) that the poloidal field is sufficiently weaker than the toroidal field, so that the displacement field discussed in the previous section is still close to being the most unstable mode.

3.5.1 Leading order estimates of the energy terms

Here, we will give estimates of the hydrostatic, toroidal, and poloidal parts of the energy in terms of the parameters of the particular displacement field constructed in the previous section. Since our R, S, and T are real functions, the cross term in equation (44) has no real part and is physically irrelevant. The total energy in terms of the integrand $\mathcal{E}$ is (from equation 27), carrying out the integration over ϕ,

$$\delta W = \delta W_{\rm hyd} + \delta W_{\rm tor} + \delta W_{\rm pol} = \frac{1}{2}\int \mathcal{E} dV = \pi \int \mathcal{E} r^2 \sin\theta dr d\theta \ . \tag{73}$$

The hydrostatic and toroidal integrands are given to leading order in $B^2/8\pi P_0$ by equations (70) and (71). On the other hand, the poloidal integrand is given by equation (44). Here, we need to make use of the azimuthal displacement, which is related to the other two components by equation (69). Since $|R| \ll |S|$ (or $\delta_r \ll \delta_\theta < 1$), to leading order we have $mT \approx 2S\cos\theta$. Thus, $R \propto 1/\delta_\theta$, $S \propto 1/\delta_r$, and $T \propto 1/\delta_r$. Each subsequent derivative ∂_x of the displacement field brings in an additional factor of $1/\delta_r$, and similarly, ∂_θ brings in a factor of $1/\delta_\theta$. It then follows that the four terms in the poloidal integrand as given by equation (44) scale as $1/\delta_r^4$, $1/\delta_r^2\delta_\theta^2$, $1/\delta_r^4$, and $1/\delta_r^2$, respectively. We thus conclude that the first and third terms in the poloidal integrand are the largest, followed by the second term, while the fourth term, which is also the only negative term in the expression, is the smallest. Thus, it becomes obvious that, for the displacement field of the form constructed here, the poloidal contribution is *overwhelmingly* positive. To leading order, keeping only the first and third terms of the poloidal integrand (equation 44), we have

$$\mathcal{E}_{\rm pol} \approx \frac{(\partial_\theta \alpha \partial_r S)^2}{8\pi r^2}\left(1 + \frac{4\cos^2\theta}{m^2}\right) \ . \tag{74}$$

We define a new coefficient for the poloidal field, in analogy to the coefficients $a_{m\neq0}$ and $c_{m\neq0}$,

$$d_{m\neq0} \equiv \frac{9}{4\pi}\left(\frac{\partial_\theta\alpha}{r^3\sin\theta}\right)^2\left(1 + \frac{4\cos^2\theta}{m^2}\right) \ . \tag{75}$$

Thus, the hydrostatic, toroidal, and poloidal energies can be written as (using equations 70 and 71, and the above definitions)

$$\begin{aligned}
\delta W_{\rm hyd} &\approx \frac{\pi}{2}\int a_{m\neq0} R^2 r^4 \sin^3\theta dr d\theta \ , \\
\delta W_{\rm tor} &\approx \frac{\pi}{2}\int c_{m\neq0} S^2 r^4 \sin^3\theta dr d\theta \ , \\
\delta W_{\rm pol} &\approx \frac{\pi}{2}\int d_{m\neq0}\left(\frac{r\partial_r S}{3}\right)^2 r^4 \sin^3\theta dr d\theta \ .
\end{aligned} \tag{76}$$

These approximations are remarkably accurate for the typical values of the parameters discussed in §3.4.7 (the errors with respect to the exact integrals are of the order of 10^{-6} or less).

In the limiting case of a vanishing area of integration (i.e. as $\delta_r \to 0$ and $\delta_\theta \to 0$ simultaneously), all slowly varying quantities can be taken as constant, and can be evaluated at the center of the displacement field (x_0, θ_0). In other words, the integrations can be carried out over the rapidly varying functions of the variables $(x - x_0)/\delta_r$ and $(\theta - \theta_0)/\delta_\theta$, and all other slowly varying functions of x and θ can be taken out of the integrations. We thus have

$$\begin{aligned}
\delta W_{\rm hyd} &\to \frac{\pi R_\star^5}{2} a_{m\neq0}(x_0,\theta_0) x_0^2 \sin^3\theta_0 \int (\partial_\theta \Pi)^2 dx d\theta \ , \\
\delta W_{\rm tor} &\to \frac{\pi R_\star^5}{2} c_{m\neq0}(x_0,\theta_0) x_0^4 \sin^3\theta_0 \int (\partial_x \Pi)^2 dx d\theta \ , \\
\delta W_{\rm pol} &\to \frac{\pi R_\star^5}{2} d_{m\neq0}(x_0,\theta_0) x_0^6 \sin^3\theta_0 \int \frac{(\partial_x^2 \Pi)^2}{9} dx d\theta \ .
\end{aligned} \tag{77}$$

The remaining integrations can be carried out in polar coordinates through the substitutions $(x - x_0)/\delta_r = \chi\cos\psi$ and

Table 1. Numerical values of the dimensionless coefficients $k_{\rm hyd}$, $k_{\rm tor}$, and $k_{\rm pol}$, defined through equation (79), for $x_0 = 0.772$ and $\theta_0 = 1.37$, corresponding to the minimum of the coefficient c_1 (Fig. 3). The background quantities are taken as in §2.2, with $p = 8\pi P_c/B_{\rm o}^2 = 10^6$, and the magnetic field structure is that of §2.3 and §2.4. We have also set $m = 1$ and $\Gamma/\gamma - 1 = 1/4$.

Coefficient	Value
$k_{\rm hyd}$	2.70
$k_{\rm tor}$	1.06
$k_{\rm pol}$	0.0832

$(\theta-\theta_0)/\delta_r = \chi\sin\psi$, yielding

$$\begin{aligned} I_\sigma &\equiv \frac{R_\star^2}{\xi_{\rm o}^2}\frac{\delta_\theta}{\delta_r}\int(\partial_\theta\Pi)^2 dx d\theta = \frac{R_\star^2}{\xi_{\rm o}^2}\frac{\delta_r}{\delta_\theta}\int(\partial_x\Pi)^2 dx d\theta = \frac{\pi\sigma}{2(\sigma-1/2)}\ , \\ J_\sigma &\equiv \frac{R_\star^2}{\xi_{\rm o}^2}\frac{\delta_r^3}{\delta_\theta}\int\frac{(\partial_x^2\Pi)^2}{9}dx d\theta = \frac{\pi\sigma^2(\sigma-1)}{6(\sigma-1/2)(\sigma-3/2)}\ . \end{aligned} \tag{78}$$

As noted following equation (72), we are interested in the case $\sigma \geqslant 2$. With these definitions, the limiting forms of the energy perturbations can be written as

$$\begin{aligned} \delta W_{\rm hyd} &\to (\Gamma/\gamma-1)P_c\xi_{\rm o}^2 R_\star k_{\rm hyd} I_\sigma\frac{\delta_r}{\delta_\theta}\ , \\ \delta W_{\rm tor} &\to -\frac{B_{\rm o}^2\xi_{\rm o}^2 R_\star}{8\pi}b_{\rm tor}^2 k_{\rm tor} I_\sigma\frac{\delta_\theta}{\delta_r}\ , \\ \delta W_{\rm pol} &\to \frac{B_{\rm o}^2\xi_{\rm o}^2 R_\star}{8\pi}b_{\rm pol}^2 k_{\rm pol} J_\sigma\frac{\delta_\theta}{\delta_r^3}\ , \end{aligned} \tag{79}$$

where $k_{\rm hyd}$, $k_{\rm tor}$ and $k_{\rm pol}$ are numerical constants which are independent of δ_r, δ_θ, and σ, and whose values are given in Table 1. The factor $\Gamma/\gamma - 1$ (also evaluated at the same point as the coefficients) has also been explicitly written in the hydrostatic part in order to keep track of the dependence on stable stratification. The amplitudes of the toroidal and poloidal fields are set by $b_{\rm tor}$ and $b_{\rm pol}$ as defined through equation (24). In addition, we have explicitly factored out all dimensional quantities $\xi_{\rm o}$, P_c, $B_{\rm o}$, and $R_\star$. The convergence of the hydrostatic, toroidal and poloidal energies to the limiting values of the approximations is demonstrated in Fig. 5.

3.5.2 Stability criteria

For stability, we must have $\delta W = \delta W_{\rm hyd} + \delta W_{\rm tor} + \delta W_{\rm pol} > 0$. Using equation (79) this can be written as, noting that $k_{\rm tor}$ is defined as a positive number and dropping common factors,

$$(\Gamma/\gamma-1)pk_{\rm hyd}I_\sigma\frac{\delta_r}{\delta_\theta} - b_{\rm tor}^2 k_{\rm tor} I_\sigma\frac{\delta_\theta}{\delta_r} + b_{\rm pol}^2 k_{\rm pol} J_\sigma \frac{\delta_\theta}{\delta_r^3} > 0\ . \tag{80}$$

Here, we have defined $p \equiv 8\pi P_c/B_{\rm o}^2$. We can rewrite this inequality as a lower bound on the amplitude of the poloidal field relative to the toroidal field,

$$\left(\frac{b_{\rm pol}}{b_{\rm tor}}\right)^2 > \delta_\theta^2\frac{I_\sigma}{J_\sigma}\left[\frac{k_{\rm tor}}{k_{\rm pol}}\left(\frac{\delta_r}{\delta_\theta}\right)^2 - \frac{k_{\rm hyd}}{k_{\rm pol}}\frac{(\Gamma/\gamma-1)p}{b_{\rm tor}^2}\left(\frac{\delta_r}{\delta_\theta}\right)^4\right]\ . \tag{81}$$

We are interested in finding the *minimum* poloidal field strength that can stabilize the magnetic field against all possible perturbations. Therefore, we would like to *maximize* the expression on the right-hand side with respect to the various parameters involved (σ, δ_r and δ_θ). First, the ratio I_σ/J_σ can be maximized with respect to σ, which for $\sigma \geqslant 2$ gives $\sigma = (3+\sqrt{3})/2 \approx 2.37$ and $I_\sigma/J_\sigma = 3(2-\sqrt{3}) \approx 0.804$ (see equation 78). Next, the expression in square brackets can be maximized with respect to the ratio δ_r/δ_θ (or, equivalently, with respect to δ_r while keeping δ_θ constant), yielding

$$\left(\frac{\delta_r}{\delta_\theta}\right)^2 = \frac{k_{\rm tor}}{2k_{\rm hyd}}\frac{b_{\rm tor}^2}{(\Gamma/\gamma-1)p}\ . \tag{82}$$

Note that for this value we indeed have $\delta W_{\rm hyd} + \delta W_{\rm tor} < 0$, as is required for the instability of the purely toroidal field in the first place. Setting $b_{\rm tor} = 1$ (which corresponds to measuring the poloidal field strength in terms of the toroidal one), $\Gamma/\gamma - 1 = 1/4$, $p = 10^6$, and using the tabulated values of the coefficients from Table 1, we obtain $\delta_r/\delta_\theta \approx 10^{-3}$.

The remaining dependence of $(b_{\rm pol}/b_{\rm tor})^2$ on δ_θ is monotonically increasing, therefore we need to evaluate the largest physically reasonable value of this parameter. For a displacement field centered at the minimum of the coefficient c_1, this value cannot be much larger than the angular extent of the negative region of the coefficient (Fig. 3). Otherwise, the toroidal energy $\delta W_{\rm tor}$ could no longer be made negative. In fact, for any sufficiently small δ_r ($\delta_r \lesssim 10^{-2}$) this condition (namely, $\delta W_{\rm tor} < 0$) translates into $\delta_\theta < 0.4$. In fact, we will take the largest value of this parameter to be about $\delta_\theta \approx 0.24$, which also

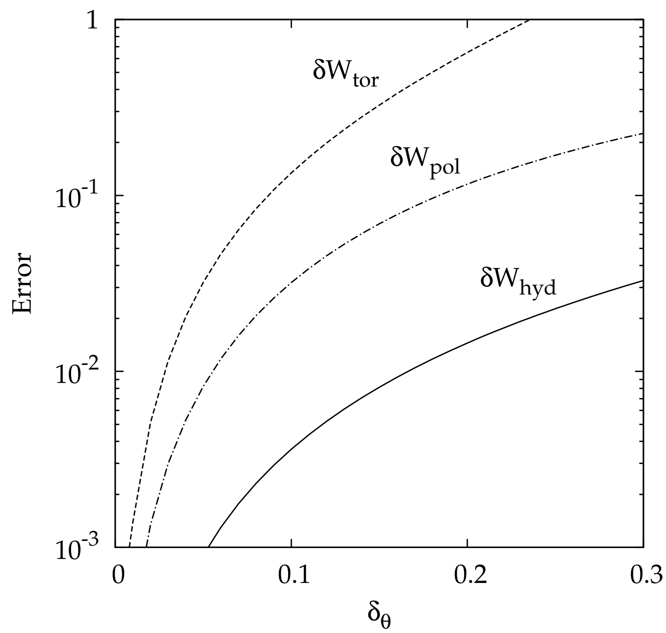

Figure 5. Errors between the limiting values (equation 79) and the exact values of the hydrostatic, toroidal and poloidal energies, as functions of δ_θ, where error = (limiting value - exact value)/exact value. All parameters, background quantities, and coefficients involved in the calculations are taken as in Table 1. In addition, δ_r and σ are taken to have the values that minimize the total energy (or, equivalently, that maximize the ratio $b_{\rm pol}/b_{\rm tor}$, as discussed in the text following equation 81), and δ_θ is allowed to vary. The exact values converge to the limiting values as δ_θ decreases.

corresponds to where our approximations start to fail (at this point the relative error between the exact and limiting values for the toroidal energy reaches 100%, corresponding to a factor of 2 error, Fig. 5).

Plugging equation (82) into equation (81), and using the values of the parameters and coefficients discussed here and in Table 1, while explicitly keeping track of the dependence on $b_{\rm tor}$, $\Gamma/\gamma - 1$ and p, we obtain

$$\left(\frac{b_{\rm pol}}{b_{\rm tor}}\right)^2 > \delta_\theta^2 \frac{I_\sigma}{J_\sigma} \frac{k_{\rm tor}^2}{4k_{\rm hyd}k_{\rm pol}} \frac{b_{\rm tor}^2}{(\Gamma/\gamma-1)p} \approx 5.8\times10^{-2}\frac{b_{\rm tor}^2}{(\Gamma/\gamma-1)p} \ . \tag{83}$$

Since $p/b_{\rm tor}^2 \gg 1$, this is a very small lower bound on the amplitude of the poloidal field needed to stabilize the toroidal field, thus effectively justifying our treatment of the poloidal field as small in comparison to the toroidal field. We can rewrite this result in terms of the energies stored in the magnetic and gravitational fields. From equations (9) and (25), we have

$$\frac{E_{\rm pol}}{E_{\rm tor}} \approx 4.3\left(\frac{b_{\rm pol}}{b_{\rm tor}}\right)^2 \qquad \text{and} \qquad \frac{E_{\rm tor}}{E_{\rm grav}} \approx 6.7\times10^{-2}\frac{b_{\rm tor}^2}{p} \ . \tag{84}$$

Replacing these in equation (83), we get

$$\frac{E_{\rm pol}}{E_{\rm tor}} \gtrsim \frac{3.7}{\Gamma/\gamma-1}\frac{E_{\rm tor}}{E_{\rm grav}} \ . \tag{85}$$

Observations provide us with an upper limit on the poloidal magnetic field strength. The above equation then gives us an upper limit on the toroidal field strength. Taking $E_{\rm pol}/E_{\rm grav} < 10^{-6}$, we get $E_{\rm tor}/E_{\rm grav} < 5.2\times10^{-4}\sqrt{\Gamma/\gamma-1}$. Thus, the maximum toroidal field strength depends on how stably stratified the star is through the factor $\Gamma/\gamma - 1$. The more stably stratified the star, the stronger the maximum toroidal field for a given poloidal field strength. For main-sequence stars, $\Gamma/\gamma - 1 \approx 1/4$ and we obtain $E_{\rm tor}/E_{\rm grav} < 2.6\times10^{-4}$, while for neutron stars, $\Gamma/\gamma-1 \sim 10^{-2}$ and we have $E_{\rm tor}/E_{\rm grav} \lesssim 5\times10^{-5}$. This also implies that a significant portion of the magnetic energy may be hidden in the toroidal field, while only the poloidal field is observed.

In particular, we can apply this result to the case of magnetars. For a $1.4M_\odot$ star with 10 km radius, the gravitational energy is $E_{\rm grav} \approx 4\times10^{53}$ erg (equation 9). The energy of the poloidal field is $E_{\rm pol} \approx 2\times10^{48}B_{15}^2$ erg (equation 25), where B_{15} is the magnetic field strength at the equator in units of 10^{15} G. (Note that, given the existence of closed poloidal field lines, this number is an order of magnitude higher than the most naive estimate obtained by multiplying the energy density corresponding to the surface field, $B^2/8\pi$, by the volume of the star.) Using $\Gamma/\gamma - 1 \sim 10^{-2}$, we then obtain an upper limit for the energy of the toroidal field as $E_{\rm tor} \lesssim 5\times10^{49}B_{15}$ erg. Note the linear dependence of the maximal toroidal energy on the surface magnetic field strength. Soft gamma repeaters (SGRs) release as much as a few 10^{46} erg energy in a single outburst (Mereghetti 2008). If the outbursts are repeated every century or so over a period of 10 millennia, then the total

energy release is of the order of a few 10^{48} erg. These could only be explained in terms of magnetars with poloidal fields in the excess of 10^{15} G. However, inclusion of a toroidal field increases the potential energy available for outbursts, and a lower surface magnetic field of the order of 10^{14} G would be sufficient to explain the observations.

A particularly interesting case is that of SGR 0418+5729, where the inferred surface magnetic field strength is just below 10^{13} G (Rea et al. 2010). Its observed X-ray luminosity is $\sim 6.2 \times 10^{31}$ erg/s, and its characteristic (spin-down) age is $\sim 2.4 \times 10^{7}$ yr. If the object is assumed to maintain the same level of activity throughout its life, then the total energy required would be $\sim 5 \times 10^{46}$ erg. Using the formulae of the preceding paragraph, the energy content of a 10^{13} G poloidal field is $E_{\rm pol} \approx 2 \times 10^{44}$ erg, i.e. more than two orders of magnitude less than what is required. On the other hand, the maximum allowed toroidal field energy in this case is $E_{\rm tor} \lesssim 5 \times 10^{47}$ erg, which would be quite sufficient.

Our results are in general agreement with the simulations of Braithwaite (2009) for a stably stratified, non-degenerate, polytropic fluid star with $\gamma = 4/3$ (i.e. a polytrope of index $n = 3$, which is a reasonable approximation for an upper main-sequence star) and $\Gamma = 5/3$. Braithwaite finds a stability condition of the form $aE_{\rm mag}/U < E_{\rm pol}/E_{\rm mag} \lesssim 0.8$, where, $E_{\rm mag} = E_{\rm pol} + E_{\rm tor}$ is the total magnetic energy and $U = E_{\rm grav}/2$ is the thermal energy (by the virial theorem). His simulations yield $a \approx 10$ for main-sequence stars, where $\Gamma/\gamma - 1 \approx 1/4$. Since in realistic stars $E_{\rm mag}/U \ll 1$, this implies that, while the poloidal component cannot be substantially stronger than the toroidal, the toroidal component can be much stronger than the poloidal.

For the lower bound on the poloidal field strength, $E_{\rm pol}$ is small and we have $E_{\rm mag} \approx E_{\rm tor}$. The condition given by equation (85) can then be written analogously to Braithwaite as $E_{\rm pol}/E_{\rm tor} \gtrsim 2aE_{\rm tor}/E_{\rm grav}$, where $a \approx 1.8/(\Gamma/\gamma - 1)$. For main-sequence stars ($\Gamma/\gamma - 1 \approx 1/4$), we obtain $a \approx 7.4$, which compares well with Braithwaite's result of $a \approx 10$. On the other hand, for neutron stars (where, we take $\Gamma/\gamma - 1 \sim 10^{-2}$, which is different than the value 1/400 used by Braithwaite), we obtain $a \sim 200$.

Notwithstanding the remarkable agreement between the two approaches, they are also notably different. First of all, while our analytic calculations can deal with arbitrary (but small) ratios of magnetic to gravitational energy, Braithwaite is forced to use a specific, and not extremely small value for his simulations ($E_{\rm mag}/U = 1/400$, compared to $\lesssim 10^{-6}$ in real stars). This means that the magnetic force is significantly stronger, and as a result, stable stratification plays a smaller role in stability. This also makes the length scale ratios in the displacement field much less extreme (see equation 82), and allows the unstable wavelengths to be resolved, even if only barely. Thus, effectively, the same instabilities should be manifested in both treatments. The proportionality of his final result to general values of the ratio $E_{\rm mag}/U$ is then stipulated. Secondly, Braithwaite's grid of values for $E_{\rm pol}/E_{\rm mag}$ has only four values for each configuration, which differ from each other by almost a factor of 2. Therefore, the value of the coefficient a is not much more precise than that. Thirdly, Braithwaite explicitly states that the coefficient c_m (for both $m = 0$ and $m = 1$) is *always* positive for the magnetic fields he considers, and that the instability of the toroidal field results entirely due to the failure of the condition $b_m^2 < 4a_m c_m$ (equation 52). The first statement seems to expressly contradict our conclusion that for *any* realistic toroidal field, c_m becomes negative at least in some regions, and is the leading source of instability, as pointed out previously by Goossens (1980) and Goossens et al. (1981) for various special field configurations. Fourthly, we do not consider the $m = 0$ case in this paper. Tayler (1973) conjectures that the $m = 1$ perturbations "seem likely to be the worst instabilities in the linear regime", and Spruit (1999) notes that the $m = 1$ mode "occurs under the widest range of conditions". However, as Braithwaite reports, there are instabilities that arise from the $m = 0$ mode as well, and need to be considered. Thus, in our treatment we have only one mode, whereas the numerical simulations in principle have all modes, so the simulations should be more unstable. It is also possible that, with the addition of the poloidal field, the particular displacement field constructed in this paper no longer corresponds to the most unstable mode, which would also imply that we are overestimating the stability. Finally, the hydrostatic background and magnetic field structure are also not identical between the two cases.

4 CONCLUDING REMARKS

Here, we summarize the main conclusions and discuss further implications of our work.

Stable stratification has an important influence on stellar magnetic equilibria and their stability, by (a) allowing a much larger assortment of possible equilibria, and (b) strongly constraining the displacement fields that might destabilize these equilibria. We can easily construct simple analytic models for axially symmetric magnetic fields compatible with hydromagnetic equilibria in stably stratified stars, with both poloidal and toroidal components of adjustable strengths, as well as the associated pressure, density and gravitational potential perturbations. This makes it possible to directly study their stability. For a weak magnetic field (in the sense that the Alfvén frequency is much smaller than the Brunt-Väisälä buoyancy frequency), the terms in the energy functional involving fluid perturbations due to the magnetic field are small and can be ignored, which makes the algebra much simpler in the cases of poloidal or mixed fields. (For a toroidal field, it does not simplify the algebra much, but it simplifies the physical interpretation.)

There is an important difference between the leading order instabilities of toroidal and poloidal fields. In toroidal fields, instabilities result from the slipping of magnetic loops around the magnetic axis (Tayler 1973). These instabilities are strongly restricted by stable stratification to surfaces of constant radius, but are not completely eliminated. As a result, a relatively weak poloidal field can be sufficient to stabilize the toroidal field (Spruit 1999). On the other hand, in poloidal fields, the instabilities result from the slipping of magnetic loops around the neutral line (Markey & Tayler 1973; Wright 1973). In this case, stable stratification is of less help in eliminating instabilities, because, while it restricts radial displacements, it does

not help with perturbations that are perpendicular to the radial direction, which are just as easily achievable in this case (ignoring curvature effects due to the fact that the neutral line is a circle). Therefore, one might expect that a relatively stronger toroidal field would be needed in order to stabilize a poloidal field. These are consistent with the upper and lower bounds found by Braithwaite (2009).

Previous literature (Tayler 1973; Goossens et al. 1981) has given proofs of instability for toroidal fields satisfying special criteria. In particular, it is often repeated that toroidal fields are unstable near the axis (Tayler 1973; Spruit 1999). Here, we prove that in fact all toroidal fields of any realistic structure, in general (barotropic and non-barotropic) fluids, are unstable. This is true even when the toroidal field is contained in a region far away from the axis, as is the case considered in this paper. The instability always happens near the high-latitude limits of the region containing the toroidal field (i.e. farthest from the equator and closest to the poles), immaterial of the exact shape of this region. This then allows us to construct a displacement field that should be a reasonable approximation for the most unstable mode, compliant with the constraints of stable stratification.

We find that the toroidal field instability considered in this paper is stabilized by a poloidal field that satisfies $E_{\rm pol}/E_{\rm tor} \gtrsim 2aE_{\rm tor}/E_{\rm grav}$ (equation 85), where $a \approx 1.8/(\Gamma/\gamma - 1)$. For main-sequence stars, we find that $a \approx 7.4$, which compares well with the factor of $a \approx 10$ obtained by Braithwaite (2009) through numerical simulations. For neutron stars, we obtain $a \sim 200$. Since observations provide us with an upper limit on the surface poloidal field, this result can then be used to place an upper limit on the internal toroidal field. We find that the energy stored in the toroidal field within the star can be significantly larger than the total energy of the poloidal field, particularly if the latter is weak. Such strong magnetic fields hidden within stars can provide a substantial additional energy budget to power magnetar activity, as well as cause significant stellar distortions with implications for precession and emission of gravitational waves (Mastrano et al. 2011).

ACKNOWLEDGEMENTS

This research was supported by FONDECYT Postdoctoral Grant 3085041, FONDECYT Regular Grants 1060644 and 1110213, CONICYT International Collaboration Grant DFG-06, a CONICYT Master's Fellowship, FONDAP Center for Astrophysics (15010003), Basal Center for Astrophysics and Associated Technologies (PFB-06/2007), Proyecto Límite VRI-PUC 15/2010, and a Melbourne University International Postgraduate Research Scholarship. We would like to thank Rafael Benguria and Andrew Melatos for useful discussions.

REFERENCES

Akgün T., Wasserman I., 2008, MNRAS, 383, 1551
Bernstein I. B., Frieman E. A., Kruskal M. D., Kulsrud R. M., 1958, Proc. R. Soc. A, 244, 17
Braithwaite J., 2009, MNRAS, 397, 763
Braithwaite J., Spruit H., 2004, Nature, 431, 819
Braithwaite J., Nordlund Å., 2006, A&A, 450, 1077
Chandrasekhar S., 1981, *"Hydrodynamic and Hydromagnetic Stability"*, Dover, New York
Chandrasekhar S., Fermi E., 1953, ApJ, 118, 116
Chandrasekhar S., Prendergast K. H., 1956, Proc. Nat. Acad. Sci., 42, 5
Dicke R. H., 1979, ApJ, 228, 898
Ferraro V. C. A., 1954, ApJ, 119, 407
Flowers E., Ruderman M. A., 1977, ApJ, 215, 302
Friedman J. L., Schutz B. F., 1978, ApJ, 221, 937
Goossens M., 1980, Geophys. Astrophys. Fluid Dynamics, 15, 123
Goossens M., Veugelen P., 1978, A&A, 70, 277
Goossens M., Biront D., 1980, SSRv, 27, 667
Goossens M., Tayler R. J., 1980, MNRAS, 193, 833
Goossens M., Biront D., Tayler R. J., 1981, Ap&SS, 75, 521
Haskell B., Samuelsson S., Glampedakis K., Andersson N., 2008, MNRAS, 385, 531
Lüst R., Schlüter A., 1954, Zs. f. Ap., 34, 263
Marchant P., Reisenegger A., Akgün T., 2011, MNRAS, 415, 2426
Markey P., Tayler R. J., 1973, MNRAS, 163, 77
Mastrano A., Melatos A., Reisenegger A., Akgün T., 2011, MNRAS, 417, 2288
Mereghetti S., 2008, A&AR, 15, 225
Mestel L., 1956, MNRAS, 116, 324
Prendergast K. H., 1956, ApJ, 123, 498
Rea N. et al., 2010, Sci, 330, 944
Reisenegger A., 2009, A&A, 499, 557
Spruit H. C., 1999, A&A, 349, 189
Tayler R. J., 1973, MNRAS, 161, 365
Wright G. A. E. 1973, MNRAS, 162, 339